

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

GABRIEL BAZO

**Desenvolvimento e análise de modelos de estatística multivariada  
baseados em Aprendizado de Máquina: predição de partos pré-termo na  
coorte de nascimento BRISA em Ribeirão Preto**

RIBEIRÃO PRETO  
2020

GABRIEL BAZO

Versão Corrigida

**Desenvolvimento e análise de modelos de estatística multivariada  
baseados em Aprendizado de Máquina: predição de partos pré-termo na  
coorte de nascimento BRISA em Ribeirão Preto**

Dissertação apresentada à Faculdade de Medicina de  
Ribeirão Preto da Universidade de São Paulo para  
obtenção do título de Mestre em Ciências.

Área de Concentração: Saúde da Criança e do  
Adolescente

Orientador: Prof. Dr. Marco Antonio Barbieri

Coorientador: Prof. Dr. Ricardo Zorzetto Nicolliello  
Vêncio

RIBEIRÃO PRETO

2020

Autorizo a reprodução total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

BAZO, Gabriel

Desenvolvimento e análise de modelos de estatística multivariada baseados em Aprendizado de Máquina: predição de partos pré-termo na coorte de nascimento BRISA em Ribeirão Preto, 2020.

126f.

Dissertação (Mestrado), Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Saúde da Criança e do Adolescente.

Orientador: Barbieri, Marco Antonio

Coorientador: Vêncio, Ricardo Zorzetto Nicolliello

1. parto pré-termo. 2. saúde perinatal. 3. políticas de saúde. 4. modelos de predição. 5. aprendizado de máquina.

Nome: BAZO, Gabriel.

Título: Desenvolvimento e análise de modelos de estatística multivariada baseados em Aprendizado de Máquina: predição de partos pré-termo na coorte de nascimento BRISA em Ribeirão Preto.

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Aprovado em:

Banca Examinadora

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Processo número 88882.378513/2019-01.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 88882.378513/2019-01.

## AGRADECIMENTOS

À minha família, minha mãe, Silvana, meu pai, Carlos e minha irmã Izabela, pelo apoio que me deram durante esta jornada e por estarem sempre presentes nos momentos de dificuldade.

À minha companheira Taíz, que me incentivou e apoiou nos momentos necessários, me conferindo forças para alcançar meus objetivos.

Ao Professor Marco Antonio Barbieri, que abriu as portas de um novo mundo, acreditou em meu potencial e assim como a água, que lentamente modela a rocha, seus conhecimentos e orientação foram dia após dia o maior agente transformador na minha jornada acadêmica.

Ao Professor Ricardo Zorzetto Nicolliello Vêncio, que desde o dia que nos conhecemos, foi um grande apoiador dessa jornada, além de todos os conhecimentos e conselhos conferidos durante a caminhada.

Para ambos os professores, Marco e Ricardo, levo comigo muito mais que apenas uma relação de professor e aluno, mas sim de amigos e exemplos acadêmicos a seguir.

Aos meus amigos discentes Paulo e Fabio, pelas incontáveis conversas, discussões e troca de conhecimentos, sem dúvida não há ciência sem filosofia.

A toda a equipe do Laboratório de Pediatria, em especial a Stela, Professora Heloísa, Conceição e Alexandre, sempre presentes em diversos momentos dessa caminhada.

Ao Departamento de Puericultura e Pediatria da Faculdade de Medicina de Ribeirão Preto da USP e a todos que direta ou indiretamente contribuíram para a realização deste trabalho.

*“The aim of science is not to open  
the door to infinite wisdom, but  
to set a limit to infinite error”*

“O objetivo da ciência não é abrir a  
porta para a sabedoria infinita,  
mas estabelecer um limite  
para o erro infinito”

Life of Galileo (1938)

**Bertolt Brecht**

## RESUMO

BAZO, G. **Desenvolvimento e análise de modelos de estatística multivariada baseados em Aprendizado de Máquina: predição de partos pré-termo na coorte de nascimento BRISA em Ribeirão Preto.** Dissertação (Mestrado). Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2020.

**Introdução:** O nascimento pré-termo (PT) é um dos principais fatores que levam a morte e causa, a longo prazo, significativa perda de capacidade, tanto física quanto psicológica entre os sobreviventes. **Objetivos:** Desenvolver uma ferramenta computacional baseada em Aprendizado de Máquina capaz de prever o nascimento PT em uma coorte de nascimentos brasileira. **Métodos:** Estudo transversal aninhado a coorte BRISA, que coletou informações sobre todos os nascimentos em 8 hospitais de Ribeirão Preto, Brasil, 2010. Foram aplicados questionários para a mãe. Nesta análise, apenas as informações maternas anteriores ao nascimento foram utilizadas. Usou-se características sociodemográficas, estilo de vida e acompanhamento pré-natal. Uma regressão foi utilizada para determinar as variáveis independentes e a variável dependente foi criada a partir da categorização da idade gestacional (<37s PT, ≥37s termo). Seis algoritmos de predição foram testados, *Random Forest* (RF), *Support vector machines* (SVM), *Naive Bayes* (NB), *Decision Tree* (DT), *Linear discriminant analysis* (LDA) e *K-nearest neighbors* (KNN). **Resultados:** 7699 gestantes participaram e 7651 atenderam todos os critérios de inclusão. A idade média foi de 27 anos ( $\pm 6,25$ ), 44,22% eram primíparas. A taxa total de PT foi de 12,7%. As variáveis com maior poder de predição para o parto PT, foram: ocorrência de um parto anterior PT, ameaça de parto PT na gestação atual, quantidade de consultas do pré-natal no segundo e terceiro trimestre da gestação, realização e tipo de serviço do pré-natal, número de fetos, hospitalização durante a gestação, ocorrência de doenças durante a gestação corrente, hipertensão gestacional, anemia e quantidade de filhos vivos. RF apresentou a maior acurácia ( $acc=0,913$ ), sensibilidade ( $s=0,81$ ) e área sob a curva ( $auc=0,842$ ). KNN apresentou a maior especificidade ( $e=0,972$ ). SVM o maior valor positivo preditivo ( $vpp=0,649$ ). NB o maior valor preditivo negativo ( $vpn=0,984$ ). Uma simulação computacional de um pré-natal, nas normas do Ministério da Saúde, em mães nas quais o modelo de predição (SVM) indicou a ocorrência de um possível parto PT, demonstrou uma redução de 3,9% no número de nascimentos PT (0,5 na porcentagem). Em âmbito nacional, no ano de 2017 seria possível uma redução de cerca de 12,7 mil partos PT, com economia média de 119 milhões de reais, para sistemas público e privado. **Conclusão:** Fatores como hospitalização durante a gravidez, hipertensão, tipo de serviço e características do pré-natal foram determinantes para a predição do parto PT. O estudo demonstrou ainda que, foi possível desenvolver modelos de predição para partos PT e sugere-se que esses modelos, adotados em nível nacional por meio de políticas públicas de saúde possam reduzir o número de partos PT, impactar o custo da saúde pública no país e melhorar a qualidade de vida da mãe e seu recém-nascido.

**Palavras-chave:** parto pré-termo, saúde perinatal, políticas de saúde, modelos de predição, aprendizado de máquina.



## ABSTRACT

BAZO, G. **Development and analysis of multivariate statistical models based on Machine Learning: prediction of preterm births in the BRISA birth cohort in Ribeirão Preto.** Dissertation (MsC). Faculty of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2020.

**Introduction:** Preterm birth (PB) is one of the main factors that lead to death and causes, in the long term, significant loss of capacity, both physical and psychological among survivors. **Objectives:** To develop a computational tool based on Machine Learning capable of predicting PB in a Brazilian birth cohort. **Methods:** This is a cross-sectional study, based on the BRISA cohort, which collected information on all births in 8 hospitals in Ribeirão Preto, Brazil in 2010. Questionnaires about the mother and the newborn were applied. In this analysis, only maternal information prior to birth was considered. Sociodemographic characteristics, lifestyle, characteristics of pregnancy and prenatal care were used. A non-parametric regression was used to determine the independent variables and the dependent variable was created from the categorization of gestational age (<37s PT, ≥37s term). Six prediction algorithms were tested, Random Forest (RF), Support vector machines (SVM), Naïve Bayes (NB), Decision Tree (DT), Linear discriminant analysis (LDA) and K-nearest neighbors (KNN). **Results:** 7699 pregnant women participated and 7651 met all inclusion criteria. The average age was 27 years old (+/- 6,25), 44.22% were primiparous. The total PB rate was 12.7%. The variables with the greatest predictive power for PB delivery were: occurrence of a previous PB delivery, threat of PB delivery in the current pregnancy, number of prenatal consultations in the second and third trimester of pregnancy, performance and type of service of the PB prenatal care, number of fetuses, hospitalization during pregnancy, occurrence of diseases during current pregnancy, gestational hypertension, anemia and number of live children. RF showed the highest accuracy (acc = 0.913), the highest sensitivity (s = 0.81) and the largest area under the curve (auc = 0.842). KNN had the highest specificity (e = 0.972). SVM had the highest positive predictive value (vpp = 0.649). NB, in turn, obtained the highest negative predictive value (vpn = 0.984). The computer simulation of prenatal care, in the Ministry of Health standards, in mothers in whom the prediction model (SVM) indicated the occurrence of a possible PB, demonstrated a 3.9% reduction in the number of PB (0.5 in percentage). Therefore, at the national level, in the year 2017 a reduction of around 12.7 thousand PB births would be possible, an average saving of 119 million reais, for both public and private systems. **Conclusion:** Factors such as hospitalization during pregnancy, hypertension, type of service and characteristics of prenatal care were determinant for the prediction of PB delivery. The study also demonstrated that it is possible to develop prediction models based on Machine Learning for PB deliveries, and it is suggested that these models, adopted at national level through public health policies, can reduce the number of PB deliveries, impacting the cost of public health in the country, and much more, the quality of life of the mother and her newborn.

**Keywords:** preterm birth, perinatal health, health policies, prediction models, machine learning.

## LISTA DE FIGURAS

<b>Figura 1:</b> Árvore de decisão e seu funcionamento .....	29
<b>Figura 2:</b> Exemplificação da margem de decisão no SVM .....	31
<b>Figura 3:</b> Modelos de kernel e suas formas distintas de classificação .....	33
<b>Figura 4:</b> Representação do funcionamento de uma Random Forest .....	34
<b>Figura 5:</b> Ilustração do Naive Bayes .....	35
<b>Figura 6:</b> Exemplificação dos K vizinhos mais próximos .....	37
<b>Figura 7:</b> Possíveis projeções para separar classes com LDA.....	39
<b>Figura 8:</b> Fluxograma do recrutamento ao nascimento e composição da população final do estudo. Ribeirão Preto, 2010.....	43
<b>Figura 9:</b> Exemplo de funcionamento do MARS com dois Splines.....	48
<b>Figura 10:</b> Validação cruzada utilizada nos modelos testados .....	50
<b>Figura 11:</b> Exemplo de matriz de confusão com suas respectivas classes .....	52

## LISTA DE TABELAS

<b>Tabela 1:</b> Variáveis obtidas a partir de feature selection .....	56
<b>Tabela 2:</b> Características das mães utilizadas na análise – Média (EP).....	57
<b>Tabela 3:</b> Características das mães utilizadas na análise - % .....	57
<b>Tabela 4:</b> Algoritmos testados no presente estudo e respectivos resultados ..	58

## LISTA DE ABREVIATURAS E SIGLAS

<b>ACC</b>	Acurácia
<b>AUC</b>	Área sob a curva
<b>BRISA</b>	<i>Brazilian Birth Cohort Studies, Ribeirão Preto and São Luís</i>
<b>CMV</b>	Citomegalovírus
<b>CRSMRP</b>	Centro de Referência da Saúde da Mulher de Ribeirão Preto
<b>CV</b>	<i>Cross validation</i>
<b>DCV</b>	Doenças Cardiovasculares
<b>DP</b>	Desvio padrão
<b>DT</b>	<i>Decision Tree</i>
<b>DUM</b>	Data da última menstruação
<b>E</b>	Especificidade
<b>ER</b>	Erro Padrão
<b>ES</b>	Sistemas Especialistas
<b>GCV</b>	<i>Generalized Cross Validation</i>
<b>HCFMRP</b>	Hospital das Clinicas da Faculdade de Medicina de Ribeirão Preto
<b>IA</b>	Inteligência Artificial
<b>IDH</b>	Índice de Desenvolvimento Humano
<b>IG</b>	Idade gestacional
<b>KNN</b>	<i>K-nearest neighbors</i>
<b>LDA</b>	<i>Linear discriminant analysis</i>
<b>MA</b>	Maranhão
<b>MARS</b>	<i>Multivariate Adaptive Regression Splines</i>
<b>NB</b>	<i>Naive Bayes</i>
<b>PN</b>	Pré-natal

<b>PT</b>	Pré-termo
<b>RF</b>	<i>Random Forest</i>
<b>RN</b>	Recém nascido
<b>RNA</b>	Redes neurais artificiais
<b>RP</b>	Ribeirão Preto
<b>RSS</b>	<i>Residual sum of squares</i>
<b>S</b>	Sensibilidade
<b>SINASC</b>	Sistema de Informações de Nascidos Vivos
<b>SIM</b>	Sistema de Informações de Mortalidade
<b>SP</b>	São Paulo
<b>SVM</b>	<i>Support vector machine</i>
<b>T</b>	Termo
<b>TCLE</b>	Termo de consentimento livre esclarecido
<b>USG</b>	Ultrassonografia
<b>VPN</b>	Valor preditivo negativo
<b>VPP</b>	Valor preditivo positivo
<b>WHO</b>	<i>World Health Organization</i>

# Sumário

<b>1 INTRODUÇÃO</b>	<b>16</b>
1.1 Objetivo	16
1.2 Estrutura da Dissertação	17
<b>2 REVISÃO BIBLIOGRÁFICA</b>	<b>18</b>
2.1 Nascimento Pré-termo	18
2.2 Desdobramentos do nascimento PT	20
2.2.1 Criança	20
2.2.2 Pais	21
2.3 Evolução da inteligência artificial e aplicações em saúde	22
2.3.1 Algoritmos de IA: passado e presente	22
2.3.2 Emprego dos algoritmos de IA em um contexto médico	24
2.3.3 Predição de partos PT: modelos clássicos e modernos	26
2.4 Como os algoritmos transformam a medicina que conhecemos hoje: desafios e perspectivas	27
2.5 Métodos de Estatística Multivariada/Aprendizado de Máquina	28
2.5.1 Decision tree (DT)	28
2.5.2 Support vector machine (SVM)	30
2.5.3 Random Forest (RF)	33
2.5.4 Naive Bayes (NB)	35
2.5.5 K-nearest neighbors (KNN)	36
2.5.6 Linear discriminant analysis (LDA)	39
<b>3 HIPÓTESE</b>	<b>41</b>
<b>4 MÉTODO</b>	<b>42</b>
4.1 Participantes	42
4.2 Instrumentos utilizados e coleta de dados	44
4.3 Criação da variável dependente	45
4.4 Análises estatísticas	45
4.4.1 Pré-processamento dos dados	45
4.4.2 Seleção dos preditores (variáveis independentes)	46
4.4.3 Desenvolvimento do modelo de predição	49
4.4.4 Treinamento	49
4.4.5 Testes e avaliação	50
4.6 Implementação dos modelos	53

4.5 Aspectos éticos .....	54
<b>5 RESULTADOS.....</b>	<b>55</b>
<b>6 DISCUSSÃO .....</b>	<b>59</b>
6.1 Limitações .....	62
6.2 Implicações na prática.....	62
<b>7 CONCLUSÃO .....</b>	<b>64</b>
<b>REFERÊNCIAS.....</b>	<b>65</b>
<b>APÊNDICE A – Tabelas complementares.....</b>	<b>73</b>
<b>APÊNDICE B – Árvore de decisão .....</b>	<b>74</b>
<b>APÊNDICE C – Códigos (R codes).....</b>	<b>75</b>
<b>APÊNDICE D – Modelo de predição categórico.....</b>	<b>79</b>
<b>APÊNDICE E – Protótipo da aplicação web (backend) .....</b>	<b>81</b>
<b>ANEXO A .....</b>	<b>86</b>
<b>ANEXO B .....</b>	<b>87</b>
<b>ANEXO C – Questionário da Mãe.....</b>	<b>88</b>
<b>ANEXO D – Questionário do RN .....</b>	<b>123</b>

## 1 INTRODUÇÃO

Segundo a Organização Mundial da Saúde, em 2018, cerca de 15 milhões de bebês nasceram pré-termo. Ou seja, mais de 1 a cada 10 bebês no mundo todo nasceu com menos de 37 semanas de idade gestacional. Deste grupo de bebês, 7% ou 1 milhão, morre decorrente a complicações do nascimento precoce (WHO, 2016).

Contudo, também de acordo com a Organização Mundial de Saúde mais de três quartos dos bebês pré-termos podem ser salvos, desde que sejam realizados cuidados durante o parto e no período pré-natal (WHO, 2018).

Por esse motivo, modelos para previsão e auxílio a decisão com foco no resultado da gestação é um assunto de crescente preocupação para profissionais de saúde e gestores públicos. Determinar quais mulheres de fato necessitam de um acompanhamento intensificado, e/ou algum tipo de intervenção, é essencial para reduzir as taxas e custos que o nascimento PT ocasiona.

Desta forma, a construção de um modelo apropriado, capaz de fazer previsões com alto nível de confiança e de forma rápida, dada a heterogeneidade de características nas mulheres gestantes tem sido o foco de muitas pesquisas recentes.

Diferentes modelos de estatística multivariada foram, portanto, desenvolvidos e comparados, a partir de dados coletados em uma coorte de nascimento, a fim de contribuir com conhecimentos e aplicações para o problema de saúde mundial contextualizado.

### 1.1 Objetivo

Geral: desenvolver uma ferramenta computacional baseada em Aprendizado de Máquina capaz de prever o nascimento pré-termo.

Específicos:

- 1 – Identificar as variáveis com maior poder de predição.
- 2 – Desenvolver um aplicativo de fácil utilização para médicos e ou profissionais de saúde.



3 – Aplicar o modelo de predição na Coorte BRISA, a fim de, identificar possíveis gestantes com risco de parto pré-termo e simular computacionalmente a realização de um pré-natal adequado de acordo com o Ministério da Saúde para posteriormente verificar se houve redução do risco.

## 1.2 Estrutura da Dissertação

Esta dissertação é composta de sete capítulos. No primeiro, Introdução, foram apresentadas as motivações do estudo e seus objetivos.

No segundo capítulo, Revisão Bibliográfica, são definidos e explicados os mecanismos biológicos que compõe um parto Pré-termo e seus desdobramentos, tanto para o recém-nascido, bem como os pais. Da mesma forma, a teoria de cada um dos métodos de Estatística multivariada aplicado no desenvolvimento dos modelos de predição de parto pré-termo.

No terceiro capítulo a hipótese é apresentada. No quarto capítulo, Métodos, são apresentados a origem e estrutura dos dados utilizados neste estudo, as técnicas utilizadas e a forma como os modelos de predição foram desenvolvidos. Neste mesmo capítulo também são definidos os indicadores de desempenho para a comparação dos modelos e a análise de *feature selection* aplicada para avaliar a importância das variáveis independentes.

No quinto e sexto capítulo, Resultados e Discussão, são apresentados a comparação do desempenho dos modelos desenvolvidos com outros relatados na literatura, bem como seus ajustes, além da análise dos indicadores utilizados para composição dos modelos e sua influência no parto pré-termo.

No sétimo e último capítulo, Conclusão, uma síntese dos resultados do estudo são apresentados junto a sugestões para próximos estudos.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Nascimento Pré-termo

Definido como parto antes de 37 semanas completas de gestação, nascer pré-termo (PT) é causado por uma etiologia multifatorial com agentes individuais e ambientais, o que torna a predição e mesmo a prevenção de um parto PT um evento desafiador durante o pré-natal (BLENCOWE et al., 2013a). Nascer PT é um dos principais fatores que causam a mortalidade neonatal, além de estar relacionado à paralisia cerebral, displasia bronco pulmonar, retinopatia da prematuridade e outras morbidades, em longo prazo, pode resultar em significativa perda de capacidade humana, tanto física quanto psicológica entre os sobreviventes (OSKOVI KAPLAN; OZGU-ERDINC, 2018).

O termo PT pode se dividir em quatro categorias de acordo com a idade gestacional (IG) no momento do parto: pré-termos tardio (*late preterm*), aqueles com IG entre 34 a 36 semanas e seis dias, pré-termos moderado (*moderately preterm*), os quais nascem com IG entre 32 a 33 semanas e seis dias, muito pré-termos (*very preterm*), nascidos entre 28 a 31 semanas e seis dias de gestação, e os pré-termos extremos (*extremely preterm*), aqueles com menos de 28 semanas de gestação (WHO, 2012). Observar a IG é importante, visto que a diminuição da mesma está associada ao aumento da mortalidade, da necessidade de cuidados neonatais e de possíveis prejuízos futuros (ROCHA, 2019).

A incidência de partos PT varia entre 5 a 10% em países desenvolvidos, alcançando 18% em países em desenvolvimento. Mais de 60% de todos esses partos concentram-se na África Subsaariana e leste asiático (LIU et al., 2016). Esse quadro representa em torno de 15 milhões de bebês PT por ano (2010) (BLENCOWE et al., 2012). Esse indicador cresce gradativamente, ocasionando mais de 1 nascimento PT a cada 10 bebês. No Brasil, segundo os dados reportados no Sistema de Informações de Nascidos Vivos (SINASC), em 2017 a razão de nascimentos PT foi de 12,5% (319.312), um aumento de 0,52% quando comparado com o ano anterior que apresentou 317.657 nascimentos PT. As regiões com maiores prevalências foram a norte (13,45%) e nordeste (12,7%),

seguida por sudeste (12,3%), sul (12,2%) e centro-oeste (11,8%) (BRASIL, 2019).

Aproximadamente, 1 milhão de mortes ocorrem devido a complicações decorrentes do nascimento PT (LIU et al., 2016). Desigualdades de renda agravam as taxas de sobrevivência, enquanto que, em países de alta renda quase todos os bebês sobrevivem, em países com baixa renda metade dos bebês nascidos com menos de 32 semanas morrem devido à falta de cuidados e recursos (WHO, 2018).

O parto PT ocorre por uma variedade de razões. A grande maioria ocorre espontaneamente, mas alguns são causados por indução precoce do parto ou até mesmo a cesariana. É importante ressaltar que, a menos que seja recomendada por um médico a cesariana não deve ser planejada para antes de 39 semanas completas (WHO, 2018). Causas comuns do parto PT incluem várias gestações, infecções, condições crônicas (diabetes e hipertensão), influência genética, entretanto, muitas vezes nenhuma causa é identificada (BLENCOWE et al., 2012). O avanço na compreensão dos mecanismos envolvidos irá promover o desenvolvimento de novas soluções para prevenir o parto PT.

Ainda, de acordo com a Organização Mundial de Saúde mais de três quartos dos bebês PT podem ser salvos, desde que sejam realizados cuidados durante o parto e no período pré-natal. Diretrizes de cuidados acerca do pré-natal devem ser obedecidas e essas incluem: intervenções chaves e aconselhamentos sobre diversos aspectos de uma gestação saudável, além de descrever efeitos prejudiciais dos usos de drogas lícitas e ilícitas durante a gestação (WHO, 2016).

A predição do parto PT pode permitir uma tomada de medidas precoce, como intervenções médicas, intensificação do acompanhamento pré-natal e até mesmo a transferência para centros de atendimento terciário e ou de referência. A identificação dos fatores de risco, e a determinação de um risco individual para cada mulher grávida, a partir de um modelo computacional preditivo têm grande importância no gerenciamento obstétrico, esse pode permitir que essas mulheres se beneficiem de estratégias de tratamentos personalizadas e ajustadas de acordo com sua necessidade (KOULLALI et al., 2016).

## 2.2 Desdobramentos do nascimento PT

### 2.2.1 Criança

Os avanços recentes no cuidado de bebês PT resultaram no aumento das taxas de sobrevivência. Contudo, o aumento de problemas comportamentais, psicológicos e motores entre os sobreviventes criaram preocupações de que esses bebês possam encontrar dificuldades em sua vida adulta (LIU et al., 2016). Diferentes estudos apontam que o parto PT, incluindo todas as categorias, pode ter consequências econômicas, médicas e sociais, as quais se manifestam de diferentes formas e em todas as fases da vida (PETROU; EDDAMA; MANGHAM, 2011).

Algumas dessas consequências não são evidentes nos primeiros dois anos de vida (HADDERS-ALGRA, 2002), portanto, é necessário um acompanhamento em longo prazo, a fim de, avaliar adequadamente algum prejuízo decorrente do parto PT. Em seus primeiros anos de vida, os bebês PT geralmente necessitam de re-hospitalização, isso pode limitar sua participação em atividades da infância, o que, pode afetar habilidades sociais (PINTO et al., 2019).

Com relação aos prejuízos, a morbidade neurológica é responsável por um risco no desenvolvimento cognitivo, motor e sensorial. Essa condição é observada em torno de 5% de todos os PT (YOUNGE et al., 2017), embora, a maioria das deficiências do desenvolvimento neurológico não seja um grande déficit, essas devem ser levadas em consideração clinicamente (HADDERS-ALGRA, 2002). As complicações neurológicas, em longo prazo, representam ainda um risco aumentado de distúrbios visuais (BLENCOWE et al., 2013b; FAROOQI et al., 2006). Miopia e hipermetropia surgem em pelo menos um quarto das crianças nascidas com menos de 28 semanas (O'CONNOR et al., 2002). A necessidade de prescrição de óculos também está relacionada com a IG, inicia-se na infância e continua na adolescência e vida adulta (SAIGAL et al., 2001).

Distúrbios no sistema imunológico podem ser responsáveis pelos altos níveis de infecções em bebês PT (GOEDICKE-FRITZ et al., 2017; HORNIK et al., 2012). O parto PT interrompe obtenção de imunidade passiva e força uma

aquisição diferente, precoce e dérmica da microbiota gastrointestinal, diferente de um bebê a termo (GROER et al., 2014).

Os déficits motores em bebês PT normalmente são identificados mais cedo, alguns deles são transitórios (hipotonia ou hipertonia e mau controle da cabeça), desaparecendo próximo dos 12 meses (PINTO et al., 2019). Alguns distúrbios psicomotores inicialmente são silenciosos, porém, diferentes estudos realizados em idade escolar (EDWARDS et al., 2011), com extremos e muito PT, encontraram um maior risco de desordem na coordenação e comprometimento motor em longo prazo.

Crianças PT, na infância, costumam apresentar sintomas de ansiedade, depressão, déficit de atenção e hiperatividade. Posteriormente, problemas com autoestima, empregabilidades e menor renda foram relatadas como frequentes na vida adulta (MATHEWSON et al., 2017).

### 2.2.2 Pais

O nascimento de uma criança PT é reconhecido como uma experiência estressante e emocionalmente exigente na vida dos pais, com possível impacto em longo prazo (CHAPIESKI; EVANKOVICH, 1997). Depressão e ansiedade foram relatadas como consequências psicológicas dos pais, e essas afetam o futuro relacionamento, entre pais e filhos, e ainda, o comportamento da criança (HOLDITCH-DAVIS et al., 2003; SINGER, 1999).

Muitos fatores são reconhecidos pelos pais como eventos estressantes, entre eles, a imaturidade, a gravidade do status médico e ainda a vulnerabilidade emocional e ou psicossocial (DUDEK-SHRIBER, 2004; MILES et al., 2002). As mães de bebês PT são descritas como mais ativas, invasivas e estimulantes na relação mãe-bebê, quando comparadas às mães de filhos nascidos termo (FORCADA-GUEX et al., 2011). Essas diferenças podem estar associadas à ansiedade materna e ao estresse traumático, principalmente se levado em consideração à influência de ambas as partes (mãe e filho) (MULLER-NIX et al., 2004).

Alguns estudos avaliaram o apego materno e suas representações em relação à gravidade da prematuridade, um potencial medo de perda e a separação pós-parto; no entanto, os resultados são subjetivos. Feldman e Weller

descobriram que as representações de apego e alterações no comportamento materno foram maiores entre as mães de bebês a termo e diminuía linearmente com a duração da separação mãe-bebê após o nascimento (FRODI; THOMPSON, 1985). Korja et al. não encontrou diferenças no apego materno entre as mães de bebês a termo e PT, porém, encontrou diferenças relacionadas à presença de sintomas de depressão materna (GOLDBERG et al., 1986).

Ainda, alguns pesquisadores mostraram ligações entre um apego reprimido e sintomas relacionados ao trauma (LIOTTI, 2004). Explorar os vínculos entre experiências maternas traumáticas e representações de apego continua sendo uma linha de pesquisa interessante, pois há evidências de que, de uma perspectiva transgeracional, as experiências traumáticas não resolvidas dos pais se relacionam ao apego da criança e, à desorganização do comportamento (MAIN; HESSE, 1990).

Forcada-Guex et al. concluiu que proporcionar aos pais oportunidades de lidar, cuidar e acompanhar o bebê PT durante o período de permanência hospitalar aumenta os sentimentos de autoconfiança. Um trabalho conjunto de pais, profissionais de saúde mental e cuidadores ajudam aos pais entenderem melhor as características e limitações do bebê PT. Isso permite superar traumas e evitar a fixação de representações negativas, que ameaçam o desenvolvimento de um relacionamento saudável entre pais e filhos.

## 2.3 Evolução da inteligência artificial e aplicações em saúde

### 2.3.1 Algoritmos de IA: passado e presente

A evolução da inteligência artificial desde sua criação até o que conhecemos hoje pode ser dividida em sete fases. Sua criação (primeira fase) remete as décadas de 1940-50, com diversos cientistas de diferentes áreas discutindo a possibilidade de criar um cérebro artificial (MCCORDUCK, 2004). A primeira disciplina acadêmica acerca do assunto foi criada em 1956. Um pouco antes em 1950, Alan Turing publicou um artigo no qual especulava sobre máquinas que pensam (TURING, 1950). Ele discutia sobre a dificuldade em definir o que é “pensar”, e com isso criou o famoso Teste de Turing, nele se uma máquina pudesse manter uma conversa por meio de um *teleprinter*, de forma

indistinguível a de um ser humano, seria razoável dizer que a máquina estava “pensando” (RUSSELL; NORVIG, 2003). Essa foi a primeira proposta séria na filosofia da IA.

Em 1956, ocorreu a Conferência de Dartmouth, considerada o nascimento da IA (KLINE, 2011). Essa foi organizada pelos cientistas Marvin Minsky (co-fundador do M.I.T) e John McCarthy (inventor da linguagem Lisp e criador do termo IA), junto aos funcionários da IBM Claude Shannon (pai da Teoria da Informação) e Nathan Rochester (inventor da linguagem Assembler e criador do IBM 701).

Os anos decorrentes após a conferência (segunda fase) foram uma era de descobertas, os programas desenvolvidos eram considerados “surpreendentes” (RUSSELL; NORVIG, 2003), computadores estavam resolvendo problemas de álgebra, provando teoremas matemáticos e aprendendo a falar inglês. Havia um otimismo advindo de todas as partes, previu-se que uma máquina totalmente inteligente seria construída em menos de 20 anos (CREVIER, 1993), grandes investimentos de agências governamentais ocorreram.

Na década de 1970 (terceira fase), a IA estava sofrendo diversas críticas e falta de financiamentos. Os pesquisadores não conseguiam resolver diversos problemas, o poder computacional limitado aliado a paradoxos matemáticos reduziu o ritmo de desenvolvimento na área, logo, o grande otimismo elevou as expectativas, quando o resultado prometido não se concretizou os financiamentos foram retirados (NATIONAL RESEARCH COUNCIL, 1999).

Nos anos 1980 (quarta fase), uma forma de programas de IA conhecida como “sistemas especialistas” (ES) (JACKSON, 1998) foram adotados por empresas e logo se tornou o foco das principais pesquisas de IA. Um ES consiste em um programa que responde ou resolve problemas sobre um domínio específico do conhecimento. Um dos primeiros a surgir foi o *Dendral* (1965) (LINDSAY et al., 1993), utilizado para leitura de espectrometria, após esse foi desenvolvido em 1972 o *Mycin* (YU et al., 1979), utilizado para diagnosticar infecções sanguíneas, esses programas mostravam novamente a viabilidade da IA. Aliado a isso essa fase ficou conhecida também como a revolução do conhecimento, devido à forma como esses sistemas eram desenvolvidos.

No final dos anos 1980 e início dos anos 1990 uma nova fase (quinta), mudou novamente os rumos da IA, com o crescimento das vendas dos computadores de mesa IBM e Apple, mais potentes e baratos que as máquinas utilizadas até então, uma indústria no valor de meio bilhão de dólares foi destruída. Mais de 300 empresas de IA encerraram suas atividades, ou foram adquiridas até o final de 1993, encerrando o primeiro ciclo comercial de IA (MORAVEC, 1994).

Já em sua sexta fase de 1993 até 2010, o campo de estudo da IA já com mais de meio século, finalmente alcançou alguns de seus objetivos antigos e estava sendo utilizada em toda indústria de tecnologia. Grande parte desse sucesso se deveu ao aumento no poder computacional com a evolução dos *hardwares*, e concentrando-se em problemas isolados. Nessa fase a IA foi fragmentada em diversos subcampos, focados em problemas ou abordagens específicas, como, a mineração de dados, robótica, reconhecimento de fala, sistemas bancários, diagnóstico médico, o mecanismo de busca do Google, entre outros (RUSSELL; NORVIG, 2003).

A IA como conhecemos hoje (sétima fase) é cercada de grandes quantidades de dados (“*big data*”) (MAGOULAS; LORICA, 2009), computadores cada vez mais acessíveis e mais rápidos e técnicas avançadas de aprendizado de máquina foram aplicadas com sucesso em diferentes problemas. Em 2016, o mercado relacionado a IA ultrapassou oito bilhões de dólares, segundo o *New York Times* novamente a IA havia atingido um “frenesi” (LOHR, 2016), novos campos de aplicação começam a surgir, e técnicas avançadas (redes neurais convolucionais e recorrentes) impulsionam pesquisas em processamento de imagens, texto e reconhecimento de fala (LECUN; BENGIO; HINTON, 2015).

### 2.3.2 Emprego dos algoritmos de IA em um contexto médico

Desde o início da IA como área de estudo propriamente dita, aplicações em saúde são desenvolvidas e aprimoradas. Como já dito anteriormente o *Mycin* (1972) (YU et al., 1979), software para diagnóstico de infecções sanguíneas, fez parte dos primórdios dessa área e muito se sucedeu após esse. Os avanços médicos e tecnológicos ocorridos permitiram o crescimento das aplicações, entre elas, uma coleta e processamento de dados mais rápidos, crescimento de banco



de dados genômico, adoção de registros de saúde eletrônicos (JHA et al., 2009), avanço no processamento de linguagem natural, análise de imagens, aumento na precisão de cirurgias assistidas por robôs (HAMET; TREMBLAY, 2017).

Diversas especialidades médicas têm se beneficiado com o aumento e aplicação de pesquisas em IA. Entre elas, a radiologia e imagens, oncologia, diagnóstico de doenças, telemedicina, prontuários eletrônicos, interação de drogas.

Na radiologia a capacidade de interpretar os resultados e analisar imagens pode ajudar os médicos a detectar pequenas alterações que um médico pode não perceber. Estudo realizado em Stanford comparou a acurácia de um algoritmo com os radiologistas no diagnóstico de pneumonia, o mesmo concluiu que o algoritmo foi mais acurado que os médicos envolvidos no estudo (RAJPURKAR et al., 2017). O câncer de pele também pode ser detectado com maior precisão, estudo publicado em 2018 mostrou que a média de precisão dos dermatologistas foi de 86,6% contra 95% de um algoritmo baseado em redes neurais (AFP, 2018).

Outras doenças já estão se beneficiando dos diagnósticos auxiliados por IA, tais como diabetes e doenças cardiovasculares (DCV), as quais estão entre as dez principais causas de morte no mundo. Elas servem de base para diversos estudos e testes, cujo objetivo é o aprimoramento diagnóstico.

Avanços nos algoritmos de processamento de linguagem natural tem permitido a realização de revisões na literatura médica já existente em busca de novas interações medicamentosas (CAI et al., 2017). Essas interações representam uma ameaça para pacientes que utilizam vários medicamentos ao mesmo tempo. Ainda, para resolver a dificuldade de rastrear todas as informações sobre medicamentos conhecidas, algoritmos de aprendizado de máquina foram criados para extrair informações sobre drogas de diversas bases literárias.

Outras aplicações como o uso da telemedicina, IA na monitoração de sinais vitais e prontuários eletrônicos permitem uma rápida comunicação e troca de informações entre médicos e outros profissionais de saúde. Esses avanços aceleram a tomada de decisão, reduzem custos médicos, aperfeiçoam tratamentos, melhoram a prevenção de doenças, e conseqüentemente,

aumentam as chances de vida dos pacientes que se beneficiam dessa tecnologia.

### 2.3.3 Predição de partos PT: modelos clássicos e modernos

A predição do parto PT é um assunto abordado pela literatura científica há mais de meio século. Em 1969, Papiernik-Berkhauer criou o primeiro score com o objetivo de prever a ocorrência de um parto PT, o qual era baseado em características da gestação, estilo de vida, status socioeconômico e histórico médico. Para cada característica um valor era atribuído e a soma dos valores indicava o risco de a gestante ter um parto PT (PAPIERNIK-BERKHAUER, 1969). Pelos próximos anos esse método foi amplamente utilizado pela literatura, até que em 1980, Creasy e colaboradores relataram que a aplicação de uma adaptação desse score em uma coorte prospectiva identificou que aproximadamente 10% dos partos seriam PT, entretanto apenas um terço do previsto pelo modelo foi, de fato, PT (CREASY; GUMMER; LIGGINS, 1980).

Posteriormente, diversos estudos exploraram a predição do parto PT, testando diferentes métodos com múltiplas abordagens baseados na estatística tradicional, tais como: coeficientes de correlações, qui-quadrado e análise de sobrevivência (GOLDENBERG et al., 1996), regressão logística (COPPER et al., 1996; MEIS et al., 1998), teste exato de Fisher e intervalos de confiança (MERCER et al., 1999).

A partir dos anos 2000 com a evolução dos métodos baseados em IA e o aumento do poder de processamento dos computadores, os estudos passaram a explorar outros métodos, nas áreas de *data mining* (mineração de dados) (CHEN et al., 2011; GOODWIN; MAHER, 2000), e *machine learning* (aprendizado de máquina) (FERGUS et al., 2013).

Dentro da área do aprendizado de máquina, de particular interesse deste estudo, diferentes modelagens já foram testadas em diversos trabalhos, tais como: *support vector machine* (VOVSHA et al., 2016), *decision trees* (HILL et al., 2008; LEE et al., 2020), *neural networks* (CATLEY et al., 2006), *naive bayes* (PAN et al., 2017), *multivariate adaptive regression splines* (MENON et al., 2014). Todos com um objetivo comum: melhorar a capacidade de predição do parto PT.

Atualmente, os trabalhos que fazem uso dos métodos citados acima utilizam-se de grandes bases de dados, normalmente advindos de estudos de coortes, tanto prospectivo, como retrospectivo. Um exemplo é o modelo de predição de parto PT desenvolvido (WATSON et al., 2017) e validado (CARTER et al., 2019), QUIPP, o qual encontra-se disponível online, em sua segunda versão, para utilização sem custos. Os criadores relataram que para mulheres com menos de 30 semanas de gestação a acurácia do modelo foi de 90,4% (CARTER et al., 2019).

#### 2.4 Como os algoritmos transformam a medicina que conhecemos hoje: desafios e perspectivas

Embora os estudos sobre as aplicações de IA na saúde busquem validar sua eficácia na melhoria dos resultados do paciente, seu uso pode introduzir novos tipos de riscos, como problemas éticos, a necessidade de regulamentação (PESAPANE et al., 2018), problemas de segurança, hackers e a falta de conhecimento acerca do funcionamento dos algoritmos pelos profissionais de saúde que o utilizam. Ainda assim, o custo-benefício das aplicações de IA na saúde é positivo e tem muito a crescer nos próximos anos.

Recentemente (fevereiro/2019), o presidente dos Estados Unidos da América assinou uma ordem, tornando o desenvolvimento e a regulação da IA uma prioridade federal. Isso demonstra a preocupação com uma evolução rápida e que medidas são necessárias para melhor utilização da mesma (METZ, 2019).

Acredita-se que a IA pode fazer muito mais do que somente melhorar diagnósticos e tratamentos, ela pode também aperfeiçoar tarefas rotineiras como tomar notas e ler exames; ajudar a reduzir custos de assistência médica, com a melhora significativa de produtividade e eficácia no fluxo de trabalho.

Entre as áreas promissoras que a IA pode avançar nos próximos anos está, o reconhecimento de padrões, com possibilidade de utilização para leitura rápida e precisa de exames médicos, slides, lesões de pele entre outros; outro avanço pode estar no desuso do teclado, com o uso de processamento de linguagem natural da fala. A IA pode, ainda, reduzir o tempo gasto em atividades repetitivas no setor médico e melhorar a qualidade do atendimento ao paciente (O'CONNOR, 2019).

## 2.5 Métodos de Estatística Multivariada/Aprendizado de Máquina

A definição exata da intersecção entre a Estatística Multivariada e Aprendizado de Máquina (*Machine Learning*) ainda é nebulosa e, portanto, neste estudo, os dois termos são usados como sinônimos, visto que, ambos abordam as técnicas aplicadas.

Os modelos multivariados são utilizados para definir uma função ou modelo que represente um fenômeno de acordo com um determinado conjunto de dados. Os dados comumente são organizados em linhas e colunas. Cada linha representa uma observação diferente, enquanto que cada coluna uma variável (variáveis independentes). A coluna final compreende as respostas do problema em questão (variável dependente).

De forma geral, os modelos são aplicados em duas situações:

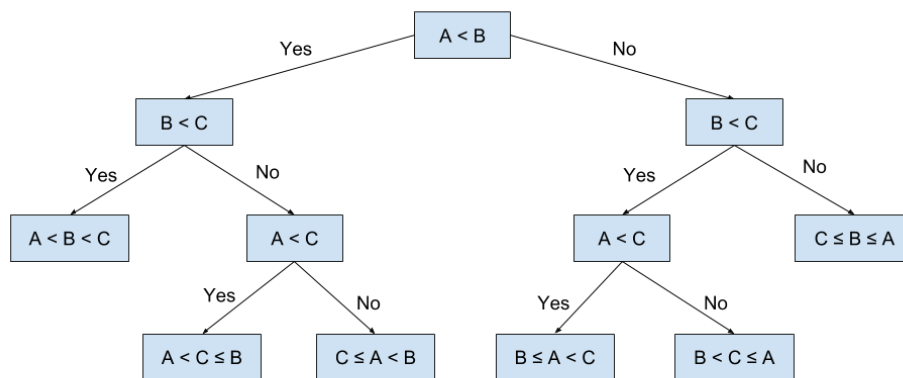
1. Regressão: obtenção de valores contínuos a partir dos dados. Neste caso a variável dependente é contínua.
2. Classificação: obtenção de uma classe respectiva para cada observação. Neste caso a variável dependente pode ser um vetor binário (0 e 1) ou tamanho  $1 \times X$ , onde  $X$  é dado pelo número de classes possíveis.

Todos os modelos utilizados no presente estudo são explicados nas subseções seguintes.

### 2.5.1 *Decision tree* (DT)

A ideia básica por trás de uma árvore de decisão (DT), é a quebra do problema em subproblemas mais simples (MURTHY, 1998), e a cada nível que descemos, menor é o grupo pertencente à aquele nó, sendo assim, recursivamente, estamos mais próximos de classificarmos aquele nó como classe A ou B (SAFAVIAN; LANDGREBE, 1991). É uma abordagem de modelagem preditiva usada em estatística, mineração de dados e aprendizado de máquina, onde essa é composta por uma estrutura tipo fluxograma na qual cada nó interno representa um “teste” em um atributo, cada ramo representa o resultado do teste e cada nó folha representa uma classe (decisão tomada após calcular todos os atributos) (QUINLAN, 1996).

Mais especificamente de acordo com (PAULA BARBOSA SOBRAL, 2003), a árvore de decisão é uma hierarquia de nós, internos ou externos, conectados por ramos. Nós internos são descritos como decisórios ou intermediários, e são esses nós onde a tomada de decisão através de teste lógico ocorre para apontar o próximo nó (filho). Nós externos são conhecidos como nós folha ou terminais, ou seja, não possuem dependente, e a eles são atribuídos uma classe.



**Figura 1:** Árvore de decisão e seu funcionamento. Fonte: (NICULAESCU, 2019)

Em geral, no sistema de classificação de uma DT, deseja-se minimizar a soma ponderada da entropia de cada folha, para que cada folha obtenha o máximo possível de uniformidade. A entropia de uma folha  $f$  é dada pela equação:

$$E_f = - \sum_{k=1}^K \left( \frac{N_{k,f}}{N_f} \right) \log \frac{N_{k,f}}{N_f},$$

onde  $N_f$  é o número de observações em uma folha e  $N_{k,f}$  é o número de observações que pertence à classe  $k$  na folha. A entropia total para uma árvore com  $F$  folhas é dada pela soma ponderada das entropias de cada folha, conforme a equação:

$$E = \sum_{f=1}^F \left( \frac{N_f}{N} \right) \cdot E_f,$$

onde  $N$  é dado pelo número de observações contidos no banco de dados de treinamento (exemplos).

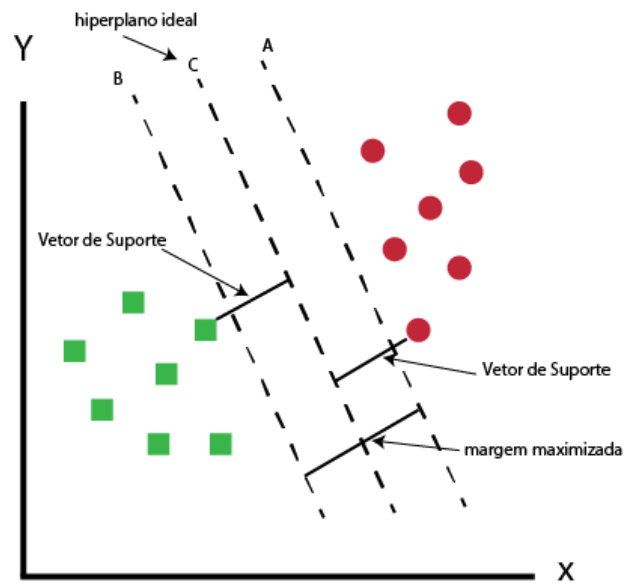
Sendo assim, em árvores de classificação, a menor entropia ocorre quando cada folha alcança o máximo de observações de mesma classe. Vale ressaltar que, quanto maior o tamanho da DT, mais ela tende a ficar super ajustada ao conjunto de dados de treinamento (exemplos), problema conhecido como *overfitting*, com isso pior será a previsão (resultado) para as observações no conjunto de dados para teste. Portanto é necessário encontrar um equilíbrio entre o tamanho da árvore e a minimização da entropia.

Em nossos testes usamos DT categóricas, onde cada nó intermediário pode se dividir de acordo com o número de categorias presente na variável independente no momento do teste lógico.

### 2.5.2 *Support vector machine (SVM)*

É definido por um hiperplano (plano em mais que duas dimensões) ou um conjunto desses, em um espaço dimensional finito ou infinito, que pode ser utilizado tanto para classificação como regressão (HEARST et al., 1998). Sua ideia fundamental é a de criar “margens”, que separem os dados de classes diferentes e que estejam afastadas o máximo possível.

Se observarmos os dados da figura 2, nota-se que as observações (pontos) possuem uma separação bem definida, ou seja, não existem observações misturadas entre as classes. Sendo assim, inúmeros hiperplanos separariam essas observações de forma eficaz. O objetivo do SVM é definir qual o hiperplano que separa os dados de forma a maximizar a distância entre eles.



**Figura 2:** Exemplificação da margem de decisão no SVM. Fonte: (CAVALCANTI, 2019)

Nesse processo de maximização, somente os dados que estão próximos dos limites das margens entre as classes serão utilizados. Esses dados (pontos) recebem o nome de “vetores de suporte”, e é por essa razão que o modelo possui o nome de *Support Vector Machine*.

Para introduzir o modelo, considera-se um caso em que as observações se dividam em duas classes apenas. As variáveis que descrevem o problema são:  $\mathbf{X}^{(n)} = (x_1^{(n)}, x_2^{(n)}, \dots, x_P^{(n)})$  com  $x \in \mathbb{R}$  e  $y^{(n)} \in \{-1, 1\}$ , sendo que  $n$  é o número de observações. O conjunto de dados ficaria conforme a tabela abaixo:

Observações	$x_1$	$x_2$	...	$x_P$	$y$
1	.	.	.	.	1
2	.	.	.	.	-1
...	.	.	.	.	-1
N	.	.	.	.	1

A equação de um hiperplano para a separação dos dados é dada na equação a seguir, onde  $\mathbf{w}$  é um vetor de tamanho  $P \times 1$  e  $b$  é um escalar:

$$\mathbf{w}' * X + b = 0$$

Dessa forma, quaisquer dados que se localize acima deste hiperplano possui  $y = 1$ , logo, qualquer dado abaixo possui  $y = -1$ . Sendo assim, definindo uma função  $g(\mathbf{X}^{(n)}) = \mathbf{w}' * \mathbf{X}^{(n)} + b$ , teremos que  $g(\mathbf{X}^{(n)}) > 0$  quando  $y^{(n)} = 1$  e, também,  $g(\mathbf{X}^{(n)}) < 0$  quando  $y^{(n)} = -1$ . Isto é, a equação a seguir é válida para todas as observações da base dados simulada neste exemplo:

$$y^{(n)}(\mathbf{w}' * \mathbf{X}^{(n)} + b) > 0 \text{ para } n = 1, 2, \dots, N.$$

Outras duas margens podem ser definidas, uma acima e outra abaixo do hiperplano que separa os dados, de acordo com as equações, nessa ordem:

$$\mathbf{w}' * \mathbf{X} + b - 1 = 0$$

$$\mathbf{w}' * \mathbf{X} + b + 1 = 0$$

Assim, como essas observações sempre estarão nas margens, acima da superior ou abaixo, a equação anterior pode ser adaptada para a seguinte forma:

$$y^{(n)}(\mathbf{w}' * \mathbf{X}^{(n)} + b) \geq 1 \text{ para } n = 1, 2, \dots, N.$$

A distância do hiperplano que separa os dados em relação à margem superior é  $\frac{1}{\|\mathbf{w}\|}$ , e em relação à margem inferior é  $\frac{-1}{\|\mathbf{w}\|}$ .

O problema abordado, em que os dados poderiam ser separados por fronteiras lineares, é trivial e pouco aplicável em problemas reais. Para generalizado o modelo para problemas não lineares, caso desse trabalho, substitui-se o produto escalar da função por Kernels.

Os Kernels mais comuns são: linear, polinomial e o radial, apresentados nas equações abaixo, respectivamente (MEYER et al, 2015):

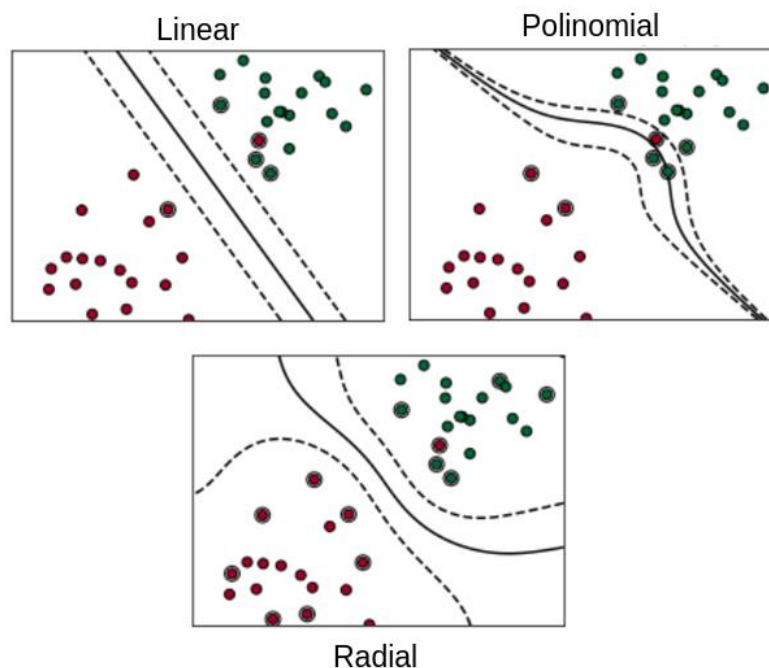
$$K(\mathbf{X}^{(n)}, \mathbf{X}^{(i)}) = (\mathbf{X}^{(n)} * \mathbf{X}^{(i)}),$$

$$K(\mathbf{X}^{(n)}, \mathbf{X}^{(i)}) = (\sigma(\mathbf{X}^{(n)} * \mathbf{X}^{(i)} + k)^d,$$

$$K(\mathbf{X}^{(n)}, \mathbf{X}^{(i)}) = \exp\left(-\sigma \|\mathbf{X}^{(n)} - \mathbf{X}^{(i)}\|^2\right).$$

A utilização dos Kernels permite que diversos problemas de classificação possam ser resolvidos, com diferentes estruturas de dados. A figura 3 apresenta exemplos de classificação com Kernel.





**Figura 3:** Modelos de kernel e suas formas distintas de classificação. Fonte: adaptado de (CHEN, 2019)

### 2.5.3 Random Forest (RF)

É um método para classificação, regressão e outras tarefas que operam a partir de várias árvores de decisão no momento do treinamento (BREIMAN, 2001). Este método se baseia no princípio que várias técnicas de aprendizagem fracas combinadas (várias árvores) podem resultar em um modelo de aprendizagem robusto (BREIMAN, 1999).

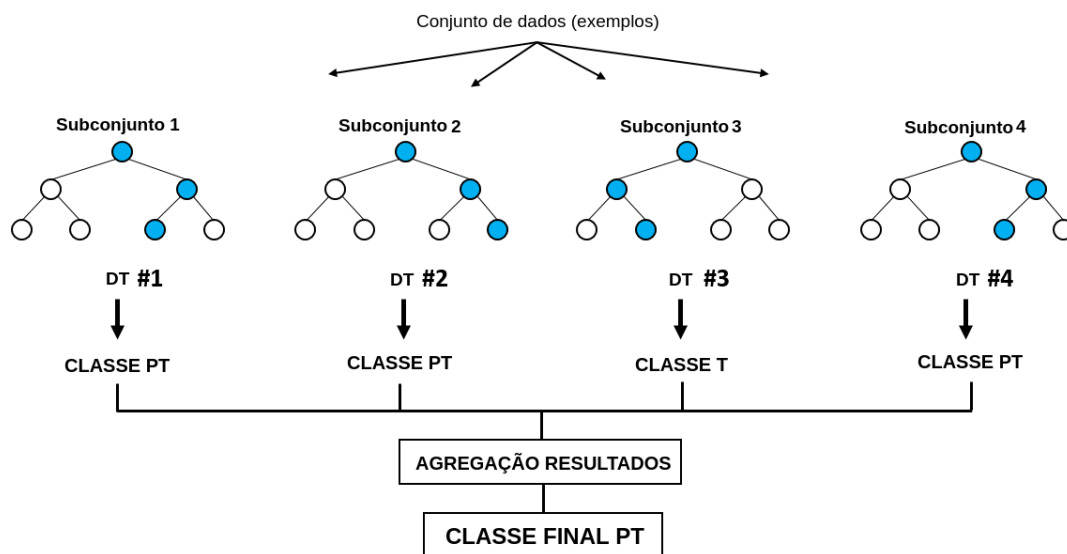
O algoritmo de RF possui duas etapas: primeiro a criação de  $B$  subsets da base original (base de treinamento com os exemplos) por amostragem com reposição (*Bootstrap Aggregation* ou *Bagging*), após isso, constrói uma DT para cada subset criado.

Na primeira etapa do algoritmo, para uma base de dados de treinamento (exemplos) com  $N$  observações, por *default* cria-se  $B$  outros subconjuntos de dados com a mesma quantidade  $N$  de observações através de amostragem aleatória com reposição, porém é possível ajustar a quantidade de subconjuntos criados. A esta etapa dá-se o nome de *bagging*. O número de subconjuntos,

acurácia do modelo e o custo computacional são linearmente proporcionais, ou seja, quanto maior o número de subconjuntos, maior a acurácia do modelo e também maior o custo computacional.

Após isso, na segunda etapa, cria-se uma DT para cada subconjunto gerado anteriormente. Estas devem ser criadas até atingir seu tamanho máximo, ou seja, até que cada folha contenha apenas uma observação (BREIMAN, 2001). Durante a criação da DT, para cada nó, um subconjunto de variáveis dentre as existentes no conjunto de dados do treinamento é selecionado aleatoriamente. A decisão de qual variável será utilizada com critério de separação do nó será realizada com base no subconjunto de variáveis selecionadas. Esse processo de seleção aleatória acontece para todos os nós da DT. A definição do número de subconjuntos gerados e o número de variáveis utilizadas em cada um é feita por conveniência e o desempenho esperado.

A figura a seguir exemplifica o que foi descrito anteriormente.



**Figura 4:** Representação do funcionamento de uma *Random Forest*. Fonte: adaptado de (ABILASH, 2018).

O resultado final é uma floresta de DT construídas de forma aleatória, em que, cada nova observação a ser classificada passa por todas às DT. Em RF diferente de uma única DT não ocorre *overfitting* (super ajuste do modelo aos dados), mesmo quando o número de DT é elevado (BREIMAN, 2001).

Finalmente para problemas de classificação, a observação a ser testada será alocada no grupo com a maior quantidade de votos.

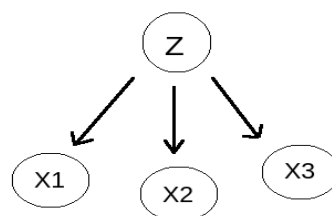
#### 2.5.4 Naive Bayes (NB)

Métodos de aprendizado Bayesianos baseiam-se em um modelo probabilístico, onde assume-se um conhecimento prévio do problema, o qual é combinado com novos exemplos de treinamento, a fim de obter-se a probabilidade final de classificação (MITCHELL, 1997).

O algoritmo *Naive Bayes* (NB) é formado por um conjunto de métodos com base no teorema de Bayes, sua principal característica. Consiste em um método baseado em probabilidades, ou seja, em um evento ocorrer dado que outro evento já ocorreu. A equação abaixo demonstra a probabilidade condicional de A dado B, em outras palavras, a probabilidade de observar A, dado que B é verdadeiro, onde A e B são eventos e  $P(B) \neq 0$  (SEGARAM, 2007).

$$P(B|A) = P(A|B) P(B) / P(A)$$

O motivo de receber o nome “*naive*” (ingênuo), se dá porque ele desconsidera completamente a correlação entre as variáveis (*features*) (ZHANG, 2004), ou seja, assume que as variáveis do conjunto de dados são independentes, isso significa que, cada variável  $x_i$  influencia uma classe  $z$ , mas não influencia outras variáveis. A figura 5 demonstra esse conceito:



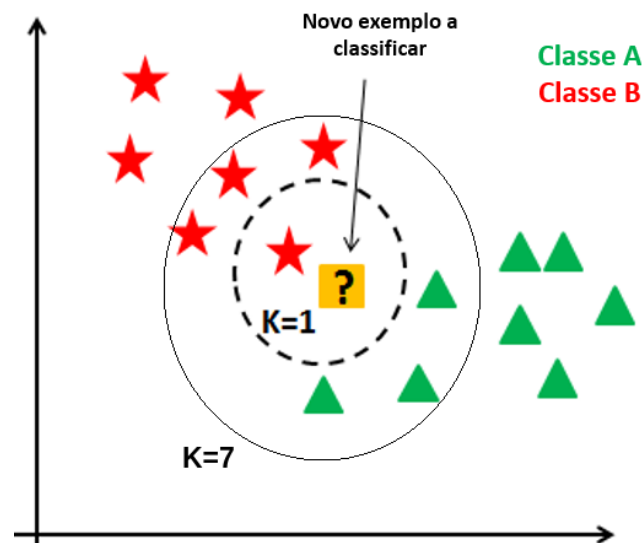
**Figura 5:** Ilustração do *Naive Bayes*. Fonte: adaptado de (Gallagher, Madden and D'Arcy, 2015).

Após isso, o classificador NB realiza o cálculo de probabilidades para a classificação através da leitura dos dados de treinamento somente uma vez para estimar as probabilidades. Ainda, NB pode assumir dois diferentes modelos estatísticos: binário (0 para não ocorrência de algum atributo ou 1 para ao menos uma ocorrência) ou multinomial (cria um vetor com o número de vezes que cada atributo ocorreu).

#### 2.5.5 *K-nearest neighbors* (KNN)

É um método não paramétrico utilizado para classificação e regressão (KELLER; GRAY; GIVENS, 1985), de maneira geral, a ideia desse algoritmo consiste em encontrar os  $k$  exemplos rotulados mais próximos do exemplo não classificado, e com base na classe desses exemplos, é tomada relativa à classe do exemplo não rotulado. Os algoritmos da família KNN possuem uma vantagem de um baixo custo computacional para realizar o treinamento do modelo. Entretanto, o custo computacional para classificar um novo exemplo pode ser alto, pois, na pior das hipóteses, esse novo exemplo deverá ser comparado com todos os exemplos contidos no conjunto de dados de treinamento.

Na figura 6 é ilustrado o funcionamento do algoritmo para um problema de classificação, com um conjunto de dados de treinamento descrito por duas classes (A e B, semelhante ao do estudo, parto T e PT). Considerando o algoritmo KNN para classificação, com  $k = 1$ , o novo exemplo (em amarelo), seria classificado de acordo com o primeiro vizinho mais próximo, que é da classe B, caso mais vizinho fossem utilizados, como o segundo exemplo onde  $k = 7$ , o novo exemplo pode assumir uma classe diferente, percebe-se que pontos da classe A passam a integrar o conjunto, porém a maioria dos vizinhos próximos ainda é predominantemente pertencentes a classe B, o que faz com que o exemplo seja classificado como B.



**Figura 6:** Exemplificação dos K vizinhos mais próximos. Fonte: adaptado de (ROCHA, 2019)

Uma questão importante é, a forma com que o grau de similaridade entre o exemplo a ser classificado e os exemplos do treinamento são definidos. Diversas medidas tem sido propostas, entre as quais estão as medidas de distância e de correlação (EVERITT, 1993).

Quando o conjunto de dados é formado por atributos numéricos, as medidas de distância podem ser aplicadas para o cálculo de similaridade entre os exemplos, sendo assim, quanto menor a distância, maior a similaridade. Índices de proximidade para o cálculo de similaridade entre os pares de exemplos devem satisfazer os seguintes pressupostos:

1.  $dist(E_i, E_j) \geq 0, \forall(i, j)$  (positividade)
2.  $dist(E_i, E_j) = 0$  se e somente se  $E_i = E_j$  (identidade)
3.  $dist(E_i, E_j) = dist(E_j, E_i)$  (simetria)

E para que um índice de proximidade seja considerado uma métrica, este por sua vez deve satisfazer, além das propriedades acima, a propriedade da desigualdade triangular:

$$4. \text{dist}(E_i, E_j) \leq \text{dist}(E_i, E_q) + \text{dist}(E_q, E_j), \forall(i, j, q)$$

Assumindo as variáveis como dimensões de um espaço multidimensional, a descrição de cada exemplo corresponde a um ponto no espaço, ou seja,  $E_i = (x_{i1}, x_{i2}, \dots, x_{in})$ . Uma maneira genérica para o cálculo das distâncias entre estes dois ou mais pontos foi estabelecido por Minkowski, de acordo com o valor do parâmetro  $d$ , que determina a medida que será utilizada.

$$\text{dist}(E_i, E_j) = \left( \sum_{l=1}^n |x_{il} - x_{jl}|^d \right)^{\frac{1}{d}}$$

Quando  $d = 1$ , a medida é conhecida como distância Manhattan, e quando  $d = 2$ , ela define a distância Euclidiana. Essa segunda é amplamente utilizada em modelos de classificação, pois possui simples implementação e grande capacidade de generalização. A distância Euclidiana entre dois pontos é dada pela raiz da soma da diferença do comprimento do segmento que conecta esses pontos elevado ao quadrado. A formula a seguir demonstra a relação descrita.

$$\begin{aligned} d(p, q) = d(q, p) &= \sqrt{((q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2)} \\ &= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}. \end{aligned}$$

Neste trabalho essa métrica foi utilizada para testes com o algoritmo KNN no desenvolvimento do modelo de predição.

Finalmente, a similaridade entre os exemplos pode também ser determinada por medidas de correlação, o qual é também medido utilizando os valores das diversas variáveis, considerando o padrão dos valores e não a magnitude. As medidas de similaridade por coeficiente de correlação são consideradas semi-métricas, pois não satisfazem a propriedade de desigualdade triangular (METZ, 2006).

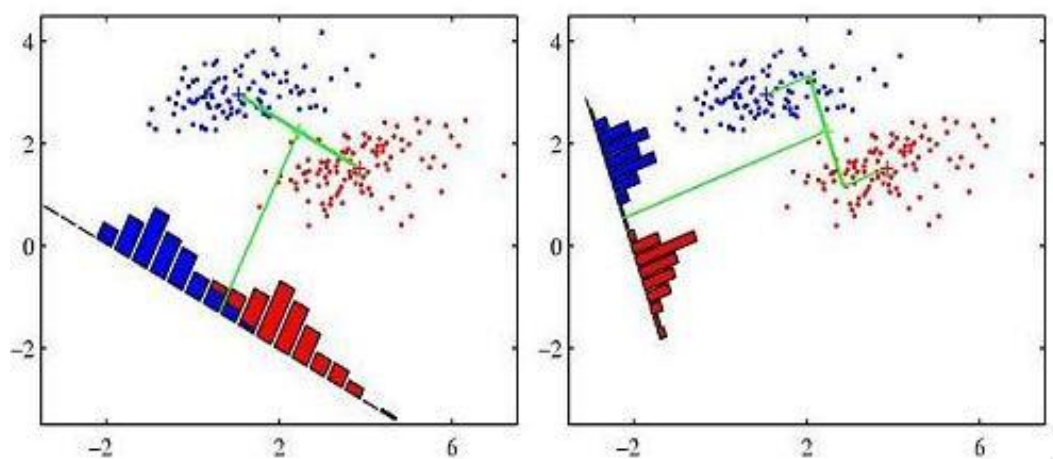
#### 2.5.6 Linear discriminant analysis (LDA)

É uma generalização do discriminante linear de Fischer, utilizado para encontrar uma combinação linear de recursos que caracteriza ou separa duas

ou mais classes de objetos ou eventos (BALAKRISHNAMA; GANAPATHIRAJU, 1998). Seu objetivo é encontrar uma combinação linear de todas as variáveis  $X_{p \times 1} = (x_1, x_2, \dots, x_p)$  de forma que o resultado dessa combinação, discriminante, tenha valores diferentes para observações pertencentes a classes diferentes e valores próximos para as observações pertencentes a mesma classe (HASTIE et al., 2008). Isto é, o discriminante  $d$  é um valor escalar resultado do produto escalar entre o vetor das variáveis  $X$  com um vetor de pesos  $a$  para cada observação, de acordo com a equação:

$$d = a \cdot X$$

Portanto, o discriminante, pode ser observado como uma projeção dos dados sobre um eixo de direção fornecido pelo vetor  $a$ , sendo este eixo obtido de tal forma que as projeções dos dados de classes diferentes possuam a maior separação possível. A figura a seguir ilustra o que foi descrito.



**Figura 7:** Possíveis projeções para separar classes com LDA. Fonte: (AWS, 2019).

Para este problema descrito, e muitos outros, pode existir mais de uma solução, isso significa que existe mais de um eixo de projeção possível para os dados, de forma que, as condições propostas sejam satisfeitas. Desta forma, o problema inicial resume-se, a um problema de autovalores e autovetores.

Sendo assim, considerando uma nova observação  $\mathbf{X}^{(0)}$  que não estava contida na base de dados de treinamento (exemplos), a alocação em um dos grupos ocorre de acordo com os passos abaixo:

1. Calculam-se todos os valores possíveis de todos os discriminantes para  $\mathbf{X}^{(0)}$ , uma para cada peso  $\mathbf{a}_d$  encontrado  $(d_1^{(0)}, d_2^{(0)}, \dots, d_D^{(0)})$

2. Calculam-se todos os valores de todos os discriminantes possíveis para as médias  $\bar{X}_k$  de todos os grupos  $k$

3. A nova observação  $\mathbf{X}^{(0)}$  é classificada no grupo onde a distância entre o discriminante da observação e o discriminante da média do grupo é a menor, de acordo com a equação:

$$\min \sum_{i=1}^D (d_i^{(0)} - \bar{d}_{i,k})^2,$$

para todos os grupos  $k = 1, 2, \dots, K$ .

O número total ( $D$ ) de discriminantes, que podem ser obtidos é igual a

$$D = \min(K - 1; P),$$

onde  $P$  é o número de variáveis independente do problema, ou seja, o tamanho do vetor  $\mathbf{X}$ .



### **3 HIPÓTESE**

Existem relações não evidentes entre as variáveis registradas na coorte BRISA que devem melhorar a performance preditiva de partos pré-termo, e que, por sua característica de interação e não-linearidade, ainda não foram descobertas pelos métodos tradicionais de Estatística descritiva.

## 4 MÉTODO

### 4.1 Participantes

Os dados utilizados neste estudo provêm de uma pesquisa de coorte intitulada “Fatores etiológicos do nascimento pré-termo e consequências dos fatores perinatais na saúde da criança: coortes de nascimentos em duas cidades brasileiras – BRISA” (acrônimo de “*Brazilian Birth Cohort Studies, Ribeirão Preto and São Luís*”), que ocorreu durante o ano de 2010 em duas cidades simultaneamente, Ribeirão Preto (SP) e São Luís (MA), esse buscou avaliar fatores associados ao nascimento pré-termo, seus desdobramentos na vida de ambas as mãe e filho e outros eventos incipientes na saúde do neonato. O estudo BRISA compreende duas coortes em cada cidade: a coorte de pré-natal (amostra de conveniência), iniciada durante a gestação, de 22 a 25 semana, e a coorte de nascimento, populacional, iniciada no momento do nascimento da criança (ROCHA, 2019). Nesse estudo para composição do modelo computacional de predição dados referentes à coorte de nascimento de Ribeirão Preto (RP) foram utilizados.

A cidade de RP está localizada na região Nordeste do Estado de São Paulo, Sudeste do Brasil. Em 2010, a população da cidade era de 604 mil habitantes, com IDH (Índice de Desenvolvimento Humano) de 0,80, ocupando na ocasião a posição de 44° cidade na classificação do IDH entre as 5565 cidades brasileiras e 19o entre as 645 cidades do Estado de São Paulo (IBGE, 2010).

A coorte de nascimentos de RP foi conduzida nos oito hospitais com serviço de maternidade da cidade, públicos ou privados, de janeiro a dezembro de 2010. Buscou-se avaliar as taxas de nascimento pré-termo e indicadores de saúde perinatal, além disso, comparando-os com os dados de coortes estudadas previamente nos últimos trinta anos em RP. Todas as puérperas procedentes do município foram convidadas a participar, sendo incluídos 7794 nascidos vivos de parto único, gemelares e natimortos (inclui as crianças do pré-natal que nasceram em 2010), com 3,8% de perdas (MORAES et al., 2018).

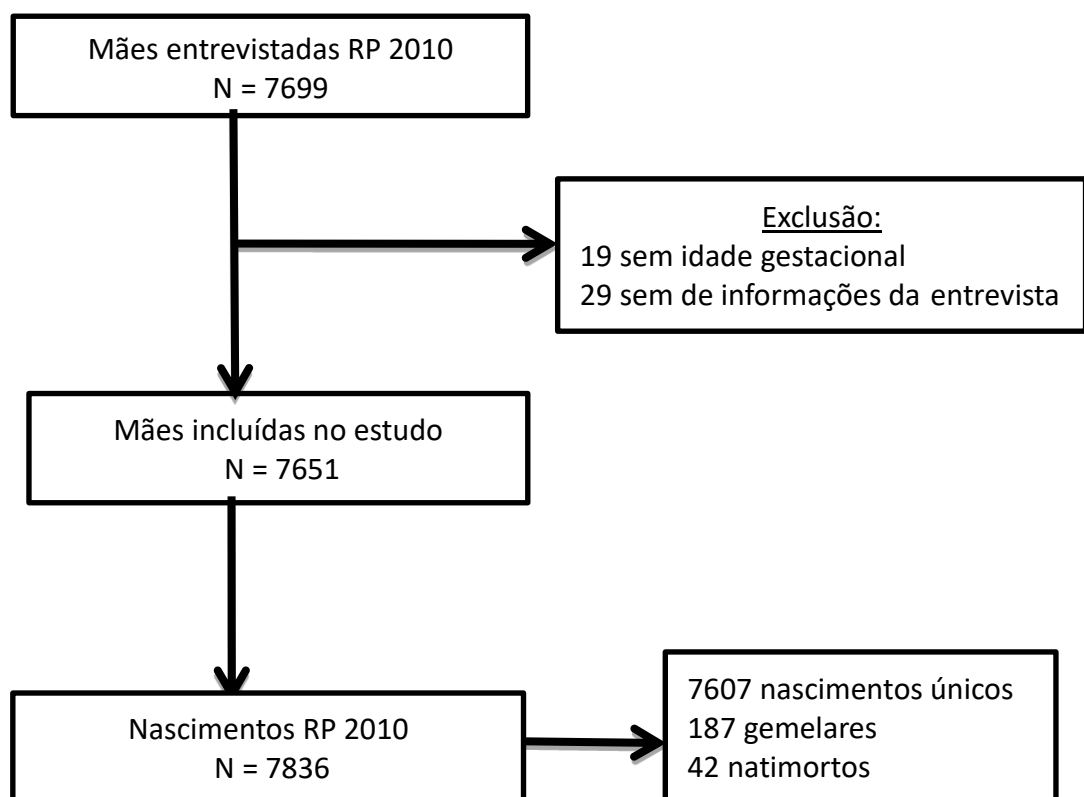
Todos os dias durante o ano de 2010, todas as maternidade de RP, com exceção aquelas com menos de 100 partos por ano, foram visitadas pelos

pesquisadores, com realização de coleta de dados através de entrevista, em casos que a mãe recebesse a alta hospitalar antes do evento da entrevista, esta era agendada e realizada em visita a domicílio (BARBOSA, 2017).

As maternidades em RP que compuseram a coleta foram: Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP/USP); Hospital São Lucas; Hospital e Maternidade Ribeirânea; Santa Casa; Centro de Referência da Saúde da Mulher de Ribeirão Preto (CRSMRP-Mater); Hospital Sinhá Junqueira e Hospital São Paulo.

No mesmo ano, durante o estudo, 98% de todos os nascimentos ocorreram em hospitais, sendo incluídos nascidos vivos de parto único e gemelares, natimortos e as crianças pertencentes à coorte do pré-natal que nasceram nesse mesmo ano. Estas duas coortes foram novamente estudadas de 13 a 30 meses após o nascimento da criança.

A figura a seguir mostra o fluxograma do recrutamento no nascimento e a composição da amostra final desse estudo.



**Figura 8** - Fluxograma do recrutamento ao nascimento e composição da população final do estudo. Ribeirão Preto, 2010

## 4.2 Instrumentos utilizados e coleta de dados

Todas as puérperas foram avaliadas por meio de entrevistas, com a utilização de questionários padronizados especificamente para o estudo, sendo um questionário para os dados da mãe e outro para o filho.

O questionário do nascimento – mãe (ANEXO C) foi composto por 11 blocos, divididos da seguinte forma:

Bloco A – Dados de identificação

Bloco B – Dados de contato

Bloco C – Dado socioeconômico e demográfico

Bloco D – Hábitos de vida

Bloco E – Dados do companheiro

Bloco F – Dados sobre saúde sexual e reprodutiva

Bloco G - Morbidades

Bloco H – Características da gestação atual e do pré-natal

Bloco I – Características do parto e do nascimento

Bloco J – Exposição ao citomegalovírus (CMV)

Bloco K – Dados do prontuário mãe

Já o questionário do nascimento do recém-nascido (RN, ANEXO D) foi dividido em dois blocos, sendo eles:

Bloco A – Dados de identificação

Bloco B – Dados do prontuário RN

Os dados obtidos através dos questionários foram duplamente digitados, a fim de permitir uma maior confiabilidade das informações. Tanto o questionário da mãe como do RN foi identificado por um código alfanumérico único, uma chave de identificação.

### 4.3 Criação da variável dependente

A variável dependente parto PT foi criada a partir da categorização da idade gestacional, binária, sendo 0 para termos e 1 pré-termo (< 37 semanas de gestação completas ou 259 dias), apesar da possibilidade de testar um modelo específico para cada categoria do PT, nesse primeiro momento testou-se o modelo com todas as categorias agrupadas.

A IG foi calculada utilizando-se a informação sobre a data da última menstruação (DUM), e a idade gestacional calculada pela USG, fornecida pela mãe no momento da entrevista. Quando havia compatibilidade entre as informações, assumindo-se erro de + 7% para USG, foi considerado o tempo de amenorreia para o cálculo da IG; caso contrário, foi considerada a informação fornecida pela USG. A idade gestacional assumida neste estudo foi ajustada para a data do parto gerando a IG definitiva (HUTCHEON et al., 2016; MEDEIROS et al., 2015).

### 4.4 Análises estatísticas

A análise estatística desse estudo se deu por etapas, desde a preparação dos dados, desenvolvimento do modelo de predição até finalmente os testes, conforme descrito a seguir; o banco de dados final contou com as informações de 7651 mães, distribuídas em 257 variáveis, do bloco A até o bloco J do questionário da mãe, o bloco K e informações sobre os recém-nascidos não foram utilizadas no desenvolvimento. Mães que não puderam ser entrevistadas ou com idade gestacional ausente foram retiradas da amostra dado critério de exclusão. Análises exploratórias foram realizadas no Stata versão 14 (College Station, Texas, USA), o modelo de predição foi desenvolvido no R versão 3.6.1.

Para caracterização da amostra foi realizada estatística descritiva. Diferenças entre os grupos PT e T foram verificadas por meio de teste qui-quadrado e teste t de *Student* não pareado.

#### 4.4.1 Pré-processamento dos dados

Inicialmente, em todos os campos que não continham informações no banco de dados (*missing values*) foi acrescentado o valor 99, esse padrão já havia sido adotado como padrão na codificação do banco de dados, isso foi necessário dado algumas características dos modelos de predição utilizados.

Algumas variáveis com a opção de não se aplica na resposta, originalmente codificada como 88, foram recodificadas para zero, por exemplo, uma mãe que informou não ter o hábito do fumo, na próxima questão quando questionada sobre a quantidade de cigarros que fez uso a resposta aplicada foi não se aplica (88), a fim de, evitar que o modelo de predição assumisse que essa consumia 88 cigarros, então esse valor foi substituído por zero. Em todas as situações semelhantes a essa, o mesmo critério foi aplicado.

Outras variáveis foram recodificadas em busca de melhora na informação, o que posteriormente resulta em melhor acurácia do modelo, tabelas 2 por 2 foram construídas para visualização de variáveis dependente e possível alteração. Por exemplo, o mês que a mãe iniciou as consultas do pré-natal, algumas responderam que não se lembrava, a questão seguinte era a quantidade de consultas por trimestre, logo se ela informasse pelo uma consulta no primeiro trimestre foi possível assumir que no mínimo o pré-natal foi iniciado no terceiro mês da gestação.

Esse processamento foi realizado para buscar uma melhor qualidade nas respostadas informadas pelas mães, visto que a coleta no nascimento foi retrospectiva e um viés de memória pode afetar a qualidade dos dados.

#### 4.4.2 Seleção dos preditores (variáveis independentes)

Com o objetivo de obter as variáveis mais importantes para compor o modelo de predição, uma análise de *feature selection* foi aplicada. A importância de uma variável é definida pela medida do efeito que alterações observadas nesta variável de exposição têm sobre a variável desfecho, ou seja, a expectativa desse efeito sobre a população. Neste modelo, todas as 257 variáveis pertinentes à pergunta do estudo foram utilizadas, somente a variável desfecho

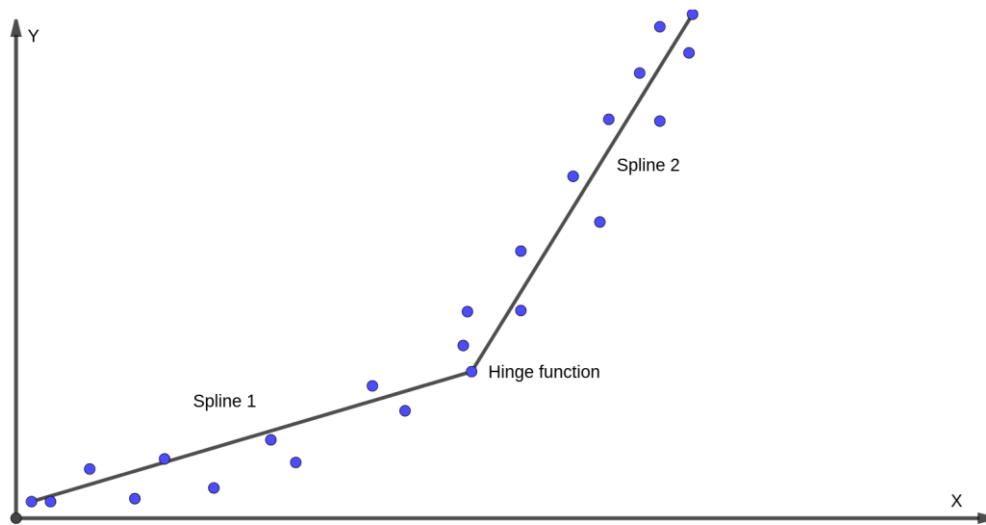
de interesse foi informada, sendo ela o nascimento PT e como resultado uma lista de variáveis foi obtida.

O modelo escolhido foi a *Multivariate Adaptive Regression Splines* (MARS) (FRIEDMAN, 1991), uma regressão não-paramétrica derivada de modelos lineares. Esse modelo divide a entrada em diferentes regiões, cada uma com sua própria equação de regressão (*Spline*) e a conexão dessas equações é feita a partir de uma função específica. Esse modelo é particularmente interessante devido à sua capacidade de solução de problemas de alta complexidade, em diversas dimensões, com variáveis contínuas ou categóricas, e de detecção e correção de variáveis altamente correlatas. Como é uma abordagem não-paramétrica, MARS não possui pressupostos sobre a distribuições das variáveis preditoras de interesse. Essa característica é importante nessa modelagem, porque cada algumas variáveis normalmente não são distribuídas.

O MARS constrói o modelo de classificação em duas etapas. Primeiro, cria um modelo excessivamente grande, adicionando funções básicas que representam transformações de variáveis únicas. Após isso, o MARS exclui as funções que contribuem menos para o modelo até que um modelo ideal seja obtido (MENON et al., 2014).

Para avaliação da lista obtida três critérios foram utilizados: o número de subconjuntos, GCV – *Generalized Cross Validation* (validação cruzada generalizada) e RSS – *Residual sum of squares* (soma dos quadrados dos resíduos). O primeiro critério, número de subconjuntos, conta quantos subconjuntos derivados do modelo principal incluem cada variável. As variáveis incluídas em uma maior quantidade de subconjuntos são consideradas mais importantes. O critério RSS calcula a diminuição na soma dos quadrados do resíduo de cada subconjunto em relação ao subconjunto anterior, para facilitar a interpretação, as reduções somadas são dimensionadas para que a maior seja 100, logo as variáveis que causam maior redução no RSS são consideradas mais importantes. Por fim, o critério GCV é calculado pelo número de variáveis presentes no subconjunto, sendo que cada variável incluída aumenta seu valor, ou seja, quanto maior seu valor, maior a importância da variável.

A figura 2 ilustra uma aplicação simples de como funciona a *Multivariate adaptive regression splines*.



**Figura 9:** Exemplo de funcionamento do MARS com dois *Splines*. Fonte: adaptado de: (CIABURRO, 2018)

Para a criação desse modelo, foi utilizado a função *earth*, presente no pacote de mesmo nome, versão 5.1. Após, a função *evimp*, contida no mesmo pacote foi utilizada para calcular a importância das variáveis.

O modelo utilizado indicou que as variáveis com maior poder de predição, dado o desfecho de interesse, foram: ocorrência de um parto anterior PT, ameaça de parto PT (antes do tempo) na gestação atual, quantidade de consultas do pré-natal no segundo e terceiro trimestre da gestação, realização do pré-natal e tipo do serviço onde o mesmo foi realizado, número de fetos, hospitalização durante a gestação, ocorrência de doenças durante a gestação atual, hipertensão gestacional, anemia e quantidade de filhos vivos (Tabela 1).

De acordo com a literatura médica, outras variáveis foram adicionadas à essas primeiras, sendo elas: hábito de fumo antes e durante a gestação, consumo de álcool na gestação, idade da mãe (SILVA et al., 2006), trabalho no início da gestação (MCDONALD et al., 1988), mês de início do pré-natal e a quantidade de ultrassons realizada em todo pré-natal (BRASIL, 2000). Essa combinação de variáveis compuseram todos os modelos desenvolvidos durante o presente estudo.



#### 4.4.3 Desenvolvimento do modelo de predição

O problema de predição (variável dependente) foi estabelecido com uma classificação binária: se a gestante irá ter ou não um filho PT. Ao todo seis algoritmos foram testados nos dados: *decision tree*, *support vector machine*, *random forest*, *naive bayes*, *knn* e *linear discriminant analysis*.

Todos os algoritmos testados pertencem à classe de algoritmos de aprendizado supervisionado (*supervised learning*), nesses é necessária uma variável de entrada e uma de saída, os mesmos “aprendem” a função de mapeamento da entrada para a saída (DUDA; STORK; HART, 2000). O objetivo é aproximar o máximo possível à função, que quando novos dados de entrada forem testados, seja possível prever a saída. Comumente esse processo é relacionado como um “professor”, com as respostas o algoritmo faz iterativamente previsões nos dados de treinamento, e é corrigido pelo “professor”, os ajustes param quando o algoritmo atinge um nível aceitável de desempenho.

Um modelo de predição categórico foi desenvolvido e testado com a finalidade de atuar de forma conjunta com o modelo de classificação binário, objetivo geral deste estudo. A finalidade do modelo categórico é complementar o resultado do primeiro modelo e tornar uma possível aplicação mais acurada (veja apêndice D).

#### 4.4.4 Treinamento

Em todos os modelos *k-folds cross validation (CV)* com  $k$  definido em cinco foi utilizado para treinamento e ajuste dos modelos de predição. Validação cruzada ou CV é uma técnica para avaliar a capacidade de generalização de um determinado modelo, a partir de um conjunto de dados (BROWNE, 2000). É amplamente empregada em problemas e modelagens de predição, se baseia no particionamento do conjunto de dados em subconjuntos mutuamente exclusivos, nesse caso cinco conjuntos, posteriormente  $k-1$  desses subconjuntos são utilizados para estimação dos parâmetros do modelo (treinamento), e um

conjunto é utilizado para validação do modelo (teste). Esse processo é realizado  $k$  vezes alternando o subconjunto de teste.



**Figura 10:** Validação cruzada (*5-folds cross-validation*) utilizada nos modelos testados.

Fonte: (OSCHINA, 2019)

#### 4.4.5 Testes e avaliação

Durante o desenvolvimento do modelo de predição, testes e ajustes foram necessários dado às características de cada algoritmo testado. Universalmente o desbalanceamento das classes da variável desfecho precisou ser ajustado para um correto funcionamento dos modelos, isso foi feito atribuindo pesos no treinamento, inversamente proporcionais à distribuição das classes, ou seja, simulando que a frequência da ocorrência foi de 50% para cada classe, esse tipo de correção é comum em modelagens computacionais quando os tamanhos das classes são assimétricos.

Para todos os modelos o *seed* (semente) foi definido com valor 123 permitindo assim a reprodutibilidade das análises. Todos os códigos utilizados no desenvolvimento e ajustes dos modelos testados são apresentados em

detalhes no apêndice B, bem como, também estão disponíveis online em: [www.github.com/gbazo/perinatalhealth](http://www.github.com/gbazo/perinatalhealth).

No SVM foi utilizado o pacote *e1071* versão 1.7-3 (MEYER et al., 2019), os dados foram normalizados através de z-score pela função “*Scale*” próprio do R (isso já ocorre por *default*), para obtenção do melhor resultado um teste com os seguintes parâmetros foi realizado: “*ranges = list(epsolon = seq(0, 1, 0.1), cost = 2^(1:6)), kernel = radial*”, esse permitiu encontrar os parâmetros que maximizam a eficiência do modelo. Em *Naive Bayes* a modelagem foi realizada com o algoritmo *default*, somente o balanceamento das classes foi aplicado, esse assim como o SVM também faz parte do pacote *e1071*.

No LDA foi utilizado a função presente no pacote *MASS* versão 7.3 (RIPLEY et al., 2019) com método estimação definido como *mle* (*maximum likelihood estimation*). No modelo com KNN função presente no pacote *caret* versão 6.0 (KUHN et al., 2020) foi utilizado a função *tune* para busca do melhor número de vizinhos próximos, o resultado foi 11, outros parâmetros foram mantidos em *default*.

Em *decision tree* os dados foram tratados antes da modelagem, para cada variável foi especificado o tipo da mesma, inteiro, numeral, binária ou categórica, isso permite uma melhora no resultado da predição, foi utilizada a função *ctree* presente no pacote *partykit* versão 1.2 (HOTHORN; SEIBOLD; ZEILEIS, 2019), uma função de poda foi aplicada na árvore, com altura definido em quatro, junto a isso a opção *multway = TRUE* foi aplicada para permitir um tratamento correto de variáveis categóricas.

Para *random forest* a função de mesmo nome presente no pacote também de mesmo nome versão 4.6 foi utilizado (LIAW; WIENER, 2018), o número de árvores definidas no modelo foi igual a 500, esse foi o ponto máximo de otimização, a partir desse valor a taxa de erro se torna linear, outros parâmetros foram mantidos em *default*.

Após o desenvolvimento dos modelos para todos uma matriz de confusão foi criada a partir da função *confusionMatrix*, parte do pacote *caret* com o atributo *positive = '1'*, isso é necessário para indicar que a classe 1 é o nascimento PT na variável.

As matrizes de confusão foram avaliadas a partir das seguintes métricas:

- Sensibilidade (taxa de verdadeiros positivos) =  $(VP / (VP + FN))$
- Especificidade (taxa de verdadeiros negativos) =  $(VN / (VN + FP))$
- Acurácia =  $(VP + VN / (VP + FN + FP + VN))$
- Valor preditivo positivo (VPP) =  $(VP / (VP + FP))$
- Valor preditivo negativo (VPN) =  $(VN / (VN + FN))$

A figura abaixo representa um exemplo teórico de uma matriz de confusão.

		PREDITO	
		Classe A	Classe B
VERDADEIRO	Classe A	VP	FN
	Classe B	FP	VN

**Figura 11:** Exemplo de matriz de confusão com suas respectivas classes. Fonte: adaptado de (ANDREONI, 2014)

## 4.6 Implementação dos modelos

Para o segundo objetivos específico do estudo, desenvolver um aplicativo de fácil utilização para médicos e ou profissionais de saúde, um protótipo de uma aplicação baseada em web (sistemas projetados para utilização através de um navegador) foi desenvolvido.

Foi utilizado o *framework Flask* como *back-end* (motor da aplicação), este é, um conjunto de classes implementadas com base em *Python*, sua escolha se deu ao fato de ser um ambiente leve, rápido e customizável o suficiente para atender as necessidades desse projeto. O *front-end* (visual da aplicação) foi desenvolvido em HTML e CSS. Os modelos desenvolvidos em R foram reescritos em linguagem *Python* com as mesmas características e parâmetros testados, isso foi necessário para permitir a compatibilidade do modelo de predição com a aplicação, além de facilitar o desenvolvimento da aplicação, visto que o R é uma linguagem voltada para modelagem estatística e científica.

Basicamente a aplicação consiste em 4 telas, uma inicial, com o título do trabalho e dois botões, um direciona para o início da predição e o segundo para uma tela com mais informações e um resumo do estudo. Na tela da predição há um formulário com as questões utilizadas no desenvolvimento do modelo de predição, após responder todas as questões um botão permite realizar a predição, o resultado da predição é exibido em uma nova tela, finalmente dessa tela por meio de dois botões é possível voltar para a tela inicial de apresentação, ou voltar para o formulário e iniciar uma nova predição.

Computacionalmente, para que o funcionamento fosse possível, foi necessário “salvar” (*dump*) o modelo desenvolvido em um arquivo específico no formato *.pkl* (*Python Pickle File*). O *back-end* da aplicação, responsável pela interação entre os elementos da aplicação, carrega o modelo de predição salvo em disco, recebe do formulário de predição as informações digitadas, realiza o teste e apresenta o resultado na próxima tela. Assim como os códigos utilizados no desenvolvimento dos modelos de predição, todo o desenvolvimento da aplicação está disponível no GitHub e no apêndice E.

#### 4.5 Aspectos éticos

O projeto temático BRISA foi aprovado pelo Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto – HCFMRP – USP (Processo nº 11157/2008) (ANEXO A). Os participantes receberam nos dias de coleta um termo de consentimento livre e esclarecido que continha informações sobre os procedimentos, seus riscos, benefícios e a garantia de confidencialidade das informações; foram entregues duas vias do termo, que foram assinados, sendo uma delas entregue ao entrevistador.

O presente também foi aprovado pelo Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP/USP) sob o parecer 3.842.040 (ANEXO B).

## 5 RESULTADOS

Do total de participantes presentes neste estudo, 87,3% foram classificados como nascidos a termo e 12,7% como PT. A comparação das características entre os grupos termo e PT (tabelas 2 e 3) foi realizada somente para as variáveis apontadas pela análise de *feature selection* (tabela 1) juntamente com aquelas selecionadas a partir da literatura.

Entre os PT 0,8% nasceu com IG < 28 semanas, 0,9% com IG entre 28 e 31 semanas e seis dias, 1,5% com IG entre 32 e 33 semanas e seis dias e 9,5% com IG entre 34 até 36 semanas e seis dias.

De todas as variáveis estudadas as únicas que não revelaram diferenças foram, a idade da mãe ( $t=-0,98$ ,  $p=0,32$ ), a média de consumo de bebidas alcoólicas no segundo trimestre da gestação ( $t=-0,29$ ,  $p=0,76$ ) e o trabalho durante a gestação ( $t=-0,12$ ,  $p=0,07$ ). Para todas as outras características diferenças entre os grupos foram observadas.

Entre as características mais relevantes, nas mães com parto PT, 61% já haviam relatado um parto PT anterior, contra apenas 7,3% das mães de T ( $p<0,001$ ) e 52% passaram pelo evento de ameaça de parto PT durante a gestação atual, contra 8,7% das mães T ( $p<0,001$ ).

A realização de acompanhamento pré-natal foi praticamente universal, com 98,7% das gestantes, nestas o tipo de serviço predominante foi público para mães termo (54,7%) e privado para mães PT (49,6%,  $p<0,001$ ). Nas mães que não realizaram pré-natal a frequência observada foi de 0,8% vs. 4,7% (T vs. PT,  $p<0,001$ ). Ainda nas mães com parto PT a primeira consulta do pré-natal ocorreu ligeiramente antes das mães T (2,4 vs. 2,5 meses,  $p=0,02$ ), a média da quantidade de consultas foi maior para mães T, tanto no segundo trimestre da gestação (2,6 vs. 2,5,  $p=0,002$ ), quanto no terceiro trimestre (4,35 vs. 2,9,  $p<0,001$ ), mães PT realizaram em média 4,4 USG contra 4,1 (PT vs. T,  $p<0,001$ ) em seu pré-natal completo. Dentre as mães que realizaram o acompanhamento do pré-natal, a recomendação médica de medicação para anemia foi maior para aquelas com parto T (65,5% vs. 58,6%,  $p<0,001$ ).

Com relação ao hábito de fumo, as mães com parto PT tiveram maior consumo de tabaco tanto antes da gestação (19,4% vs. 25,4%,  $p<0,001$ ), quanto

durante a gestação (11% vs. 16,9%,  $p < 0,001$ ), quando comparadas às mães de nascidos T.

Em adição, a necessidade de hospitalização durante a gestação em curso foi três vezes maior nas mães com parto PT (PT=21,4% vs. T=7,45%,  $p < 0,001$ ). Dentre as morbidades estudadas, a hipertensão gestacional foi a única entre as variáveis com maior poder preditivo para um parto PT, sendo que esta morbidade ocorreu duas vezes mais em mães com parto PT (PT=22,9% vs T=11%,  $p < 0,001$ ) e de maneira geral, a variável “outras doenças” ocorridas durante a gestação também foi mais frequentes em mães com parto PT (PT=22,5% vs. T=18,6%,  $p < 0,001$ ), quando comparadas às mães de nascidos T.

A tabela 1 apresenta os resultados obtidos pela análise de *feature selection*. As variáveis contidas abaixo são as que possuem o maior poder de predição para o parto PT e foram utilizadas no desenvolvimento do modelo computacional.

**Tabela 1: Variáveis obtidas a partir de feature selection**

Variáveis	Subconjuntos	GCV *	RSS**
Parto anterior PT	19	100	100
Ameaça de parto PT	18	57,8	58,6
Número de fetos	17	41,8	43,2
Qtd de consultas PN 3 TRI	15	34,8	36,3
Tipo de serviço PN	13	28,4	30,1
Necessitou de hospitalização	12	25,4	27,2
Doenças durante a gestação	9	18,9	20,7
Número de filhos vivos	7	15,3	17,0
Parou de trabalhar	5	11,5	13,2
Hipertensão gestacional	4	9,5	11,1
Qtd de consultas PN 2 TRI	3	7,1	8,7

\*GCV – *Generalized cross validation* (validação cruzada generalizada) \*\*RSS – *Residual sum of squares* (soma dos quadrados dos resíduos).

As tabelas 2 e 3 apresentam as características das mães incluídas nos grupos, acompanhadas de respectivas médias ou frequências e valores p.

**Tabela 2: Características das mães utilizadas na análise – Média (EP)**

Variáveis	Termos	Pré – termos	p-Valor*
-----------	--------	--------------	----------



Idade da mãe (anos)	27 (0,07)	27,2 (0,2)	0,32
Consultas PN 3 TRI	4,35 (0,02)	2,9 (0,06)	< 0,001
Consultas PN 2 TRI	2,6 (0,01)	2,5 (0,04)	0,002
Média consumo álcool 2 TRI (g)	2 (1,2)	3 (1,8)	0,76
Filhos vivos	1,8 (0,03)	3 (0,32)	< 0,001
Número de fetos	1 (0,0008)	1,1 (0,008)	< 0,001
Mês início PN	2,5 (0,01)	2,4 (0,04)	0,02
Quantidade de USG	4,1 (0,02)	4,4 (0,06)	< 0,001
Semanas trabalhadas gestando	2,2 (0,09)	2,7 (0,34)	0,07

\* Teste t de *Student* não pareado

**Tabela 3: Características das mães utilizadas na análise - %**

Variáveis	Termos	Pré-termos	p-Valor*
<b>Parto anterior PT</b>			< 0,001
Sim	7,3	61	
Não	91,4	37,4	
NA	1,3	1,6	
<b>Ameaça de parto PT na gestação atual</b>			< 0,001
Sim	8,7	52	
Não	91,1	46,4	
NA	0,2	1,6	
<b>Tipo de serviço PN</b>			< 0,001
Público	54,7	45,3	
Privado	44,3	49,6	
Sem PN	0,8	4,7	
NA	0,2	0,4	
<b>Realização do PN</b>			< 0,001
Sim	99,2	95,3	
Não	0,8	4,7	
<b>Hospitalização durante a gestação</b>			< 0,001
Sim	7,45	21,4	
Não	92,5	78,4	
NA	0,05	0,5	
<b>Hipertensão gestacional</b>			< 0,001
Sim	11	22,9	
Não	88,8	76,4	
NA	0,2	0,7	
<b>Habito de fumo passado</b>			< 0,001
Sim	19,4	25,4	
Não	80,6	74,5	
NA	0	0,1	
<b>Fumou durante a gestação</b>			< 0,001
Sim	11	16,9	
Não	89	83	
NA	0	0,1	
<b>Apresentou doenças durante a gestação</b>			< 0,001
Sim	18,6	22,5	
Não	81,2	77	
NA	0,2	0,5	
<b>Foi recomendado medicamento para anemia</b>			< 0,001
Sim	65,5	58,6	
Não	33,5	35,9	
Sem PN	0,8	4,7	
NA	0,2	0,8	

\* teste do qui-quadrado

A tabela 4 apresenta os resultados de cada modelo de predição testado neste estudo.

**Tabela 4: Algoritmos testados no presente estudo e respectivos resultados**

<b>Modelo</b>	<b>ACC</b>	<b>S</b>	<b>E</b>	<b>VPP</b>	<b>VPN</b>	<b>AUC</b>
<i>SVM</i>	0,824	0,422	0,936	0,649	0,853	0,818
<i>LDA</i>	0,868	0,615	0,879	0,184	0,980	0,804
<i>Random Forest</i>	0,913	0,813	0,922	0,502	0,980	0,842
<i>KNN</i>	0,879	0,237	0,972	0,554	0,897	0,511
<i>Naive Bayes</i>	0,862	0,565	0,871	0,119	0,984	0,741
<i>Decision Tree</i>	0,880	0,502	0,937	0,548	0,925	0,835

ACC – Acurácia; S – Sensibilidade; E – Especificidade; VPP – Valor preditivo positivo; VPN – Valor preditivo negativo; AUC – Área sob a curva.

Cada modelo testado apresentou características diferentes entre as métricas de avaliação. O modelo mais acurado foi o *Random Forest* (acc=0,913), o qual também obteve a maior sensibilidade (s=0,81) e a maior área sob a curva (auc=0,842). O *Knn* apresentou a maior especificidade (e=0,972). O *SVM* obteve o maior valor positivo preditivo (vpp=0,649) e o maior valor preditivo negativo resultou do *Naive Bayes* (vpn=0,984). O *Decision tree* e o *LDA* apresentaram resultados intermediários entre todos os testados. As tabelas complementares com os dados da performance de cada modelo (matriz de confusão com métricas) estão presentes no apêndice A.

## 6 DISCUSSÃO

O presente estudo teve como objetivo desenvolver uma ferramenta computacional baseada em Aprendizado de Máquina, capaz de prever o nascimento pré-termo, além de, identificar as variáveis com maior poder de predição e simular computacionalmente sua aplicação. Para tanto, utilizou algoritmos já testados em outros trabalhos (LEE et al., 2020; MENON et al., 2014; VOVSHA et al., 2016) , porém com uso de dados proveniente de um estudo de coorte. A partir dos resultados da análise de *feature selection* foi possível desenvolver e testar o desempenho dos algoritmos.

Em geral, o gerenciamento das pacientes por obstetras é baseado na estimativa de risco, aconselhamentos e na tomada de decisão. Entretanto, os métodos de estimativa de risco tradicionais e comumente utilizados aplicam o mesmo nível de risco para todas as pacientes. A fim de obter resultados mais precisos e individualizados, ferramentas preditivas e prognósticas com base em modelos estatísticos e conceitos de Aprendizagem de Máquina, demonstraram melhor julgamento clínico para prever a probabilidade dos resultados (LEE et al., 2020). A primeira tentativa de aplicação de uma ferramenta preditiva em um ambiente obstétrico não pode ser individualizada e seu resultado foi impreciso (PAPIERNIK; KAMINSKI, 1974).

Alguns estudos, diferente desse, buscaram realizar a predição com conjuntos de informações específicas sobre a gestação, tais como: eletromiografia uterina (BEDEEUZZAMAN; ANJU, 2019; SHANIBA ASMI; SUBRAMANIAM; IQBAL, 2019), comprimento cervical, fibronectina fetal (SON; MILLER, 2017), fluido cervical (OSKOVI KAPLAN; OZGU-ERDINC, 2018), análises de proteínas (GUNKO; POGORELOVA; LINDE, 2016), fluido amniótico (BARALDI et al., 2016), microRNA (ELOVITZ et al., 2015). Resultados variados foram observados, desde modelos de predição acurados, quando utilizado eletromiografia uterina, até a descoberta de ausência de relação entre o objeto de estudo e o parto PT, onde a análise dos microRNAs de mães PT não apresentou diferença significativa quando comparado a mães T.

Outros estudos, assim como esse, utilizaram de características socioeconômicas, estilo de vida, histórico médico, nível de educação e

características do PN (KOULLALI et al., 2016; LEE et al., 2020; PAN et al., 2017), ou seja, informações que descrevem o espectro gestacional de maneira ampla.

A utilização de fatores como estilo de vida, características do pré-natal, e morbidades maternas no desenvolvimento do modelo de predição apresenta vantagens e desvantagens quando comparada a outros tipos de dados utilizados. A maior vantagem está relacionada com o baixo custo para obtenção dos dados, visto que, os mesmos compõem a rotina e acompanhamento da gestante, além de que, essas informações são de fácil acesso e utilização para gestores de saúde, por meio dos dados públicos. Suas desvantagens refletem principalmente em uma redução da acurácia quando comparados com modelos construídos a partir de dados específicos, como por exemplo dados de sinais de eletromiografia uterina (99,8% vs 91,3%), quando comparado ao modelo de predição baseado em RF que obteve a maior acurácia (SHANIBA ASMI; SUBRAMANIAM; IQBAL, 2019).

O resultado da análise de *feature selection* indicou que, o fator mais importante para predição do parto PT, de acordo com os dados estudados, foi a ocorrência de um parto PT prévio (gcv=100,rss=100), e de fato 55% das mulheres que tiveram um parto PT anterior, tiveram um segundo parto PT (PT=55% vs T=45%,  $p<0,001$ ).

Para o segundo fator mais importante, mulheres que apresentaram uma ameaça de parto PT em qualquer momento da gestação (gcv=57,8; rss=58,6), 46,6% resultou em um parto PT (PT=46,6% vs. T=53,4%,  $p<0,001$ ).

Quando combinados, 78,7% das mulheres que apresentaram as duas características, apresentaram um parto PT (PT=78,7% vs. T=21,3%,  $p<0,001$ ). Somente esses dois fatores, individualmente e principalmente quando associados já seriam suficientes para uma possível intensificação no acompanhamento PN.

Já é bem estabelecido que a realização de um PN adequado é importante, para prevenção de diversos fatores, muito além do parto PT. (ICKOVICS, 2003).

Outras variáveis, como mês de início do acompanhamento do PN, quantidade de consultas, número de USGs, doenças durante a gestação, parecem estar relacionadas com a adequação do PN, ou seja, a realização de um acompanhamento que seja suficiente dado as necessidades da gestante.

Quanto ao local de realização do PN, em que, o serviço privado apresenta mais partos PT (49,6%) quando comparado ao público (45,3%), isso pode ser devido a fatores não médicos que foram mais associados com os partos operatórios do que fatores biológicos ou clínicos das gestantes. (ALMEIDA et al., 2008).

A análise de *feature selection* deixou claro ainda, uma possibilidade de direcionamento metodológico dos questionários de coleta, de duas formas, primeiro, com a determinação das variáveis independentes que melhor respondem a variável dependente, um delineamento na forma com que estas são utilizadas pode resultar em uma melhor acurácia do questionário (realizar ajustes nas perguntas), segundo, em determinadas situações, estudos com uma quantidade finita de recursos podem encontrar dificuldades para otimizar sua execução. Com a determinação das variáveis que melhor respondem uma pergunta de estudo, métodos de coleta de informações podem ser otimizados (reduzir a quantidade de perguntas e ainda assim alcançar o mesmo resultado), isso gera economia de recursos para o pesquisador, diminuição no tempo gasto em entrevistas e uma possibilidade de aumento na quantidade de participantes coletados.

É importante ressaltar ainda que, as afirmações acima não são verdades absolutas, porque para que o possível direcionamento metodológico possa ocorrer, grandes estudos, com coletas de dados que abrangem um amplo espectro sobre um assunto são essenciais.

A principal força deste estudo está ligada ao fato que, os modelos para predição do parto PT são baseados em critérios amplamente utilizados e em uma combinação de fatores de risco conhecidos, obtidos por meio de questionários sobre sintomas e outras características subjetivos. Outro ponto forte do estudo é que todas as variáveis utilizadas nos nossos modelos preditivos se baseiam em informações de fácil avaliação dos pacientes.

A abordagem mais eficaz para prevenir o parto prematuro é baseada na história obstétrica individual, o que torna a identificação de mulheres com risco para um parto PT uma tarefa importante (LEE et al., 2020).

## 6.1 Limitações

Os dados utilizados no desenvolvimento dos modelos de predição são provenientes de um estudo de coorte prospectivo, mas que no momento da coleta fez perguntas sobre eventos ocorridos no passado, estas questões são sujeitas a um viés de memória, a puérpera entrevistada podia ter dificuldades para se lembrar de ocorridos durante seu acompanhamento pré-natal e ou hábitos de vida que antecedem sua gestação. Para tanto, sempre que necessário informações do banco de dados foram revisadas nas fichas originais e até mesmo nos prontuários hospitalares disponíveis no momento do estudo.

Apesar das limitações encontradas, este estudo desenvolveu modelos de previsão personalizados de risco de nascimento pré-termo usando uma ampla gama de variáveis sobre as características da gestante obtidas após o nascimento da criança. Portanto, esses modelos podem auxiliar clínicos na tomada de decisões apropriadas para hospitalização ou treinamento de estilo de vida em ambiente ambulatorial possam ser tomadas.

## 6.2 Implicações na prática

Com os modelos de predição desenvolvidos, houve por parte dos pesquisadores envolvidos no estudo um desejo de testá-los em um serviço de PN, a fim de, validar seu funcionamento, porém isso não foi possível em tempo hábil. Para contornar essa questão, simulações computacionais foram realizadas com a mesma base de dados utilizada no desenvolvimento dos modelos.

A simulação ocorreu da seguinte forma: foi utilizada a mesma base de teste criada durante o processo de *cross validation*, o modelo de predição escolhido foi o SVM, esta escolha foi devido ao seu VPP. Neste conjunto de dados, assumiu-se um pré-natal de acordo com as normas do Ministério da Saúde (foi levado em consideração mês de início do PN, quantidade de consultas e de exames de USG), em mães nas quais o modelo de predição indicou a ocorrência de um possível parto PT, após isso, a base de teste foi submetida novamente no modelo de predição, o resultado demonstrou uma redução de 3,9% no número de nascimentos PT (0.5 na porcentagem), ou seja,

se o PN conforme o Ministério da Saúde recomenda fosse respeitado, a taxa de nascimentos PT poderia ser reduzida 0.5 pontos percentuais.

Quando aplicamos esses números em âmbito nacional, de acordo com o SINASC no ano de 2017, ocorreu no Brasil 319.312,00 mil nascimentos vivos de crianças PT, sem levar em conta que este número pode ser maior se considerarmos partos PT que não resultaram em nascimentos vivos, notificados em outros sistemas, como SIM (Sistema de Informação de Mortalidade), com uma redução de 0.5 pontos percentuais no observado de 2017, seria possível reduzir 12.773 mil nascimentos PT.

Indo mais além, isso representaria uma economia média de até 119 milhões de reais, para ambos sistemas, público e privado, tendo como base o custo médio de permanência hospitalar de US\$ 2.323,00 (MWAMAKAMBA; ZUCCHI, 2014).

## 7 CONCLUSÃO

Fatores como hospitalização durante a gravidez, hipertensão, tipo de serviço e características do pré-natal foram determinantes para a predição do parto PT. O estudo demonstrou ainda que, foi possível desenvolver modelos de predição baseados em Aprendizado de Máquina para partos PT, e que estes possuem capacidade de aplicação em cenários reais. Para definir qual técnica de Aprendizado de Máquina obteve o melhor resultado, deve-se levar em consideração as características da população e do teste (maior sensibilidade ou especificidade), contudo para a população deste estudo os modelos baseados em *Random Forest* e *Support Vector Machines* obtiveram os melhores resultados.

Sugere-se ainda que os modelos, adotados em nível nacional ou até mesmo em países com características semelhantes, por meio de políticas públicas de saúde possam impactar o custo da saúde pública, e muito além disso, a qualidade de vida da mãe e seu recém-nascido.



## REFERÊNCIAS

- ABILASH, R. **Applying Random Forest (Classification) — Machine Learning Algorithm From Scratch With Real Datasets**. Disponível em: <<https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57>>. Acesso em: 15 jan. 2020.
- AFP. Computer learns to detect skin cancer more accurately than doctors. **The Guardian**, 2018.
- ALMEIDA, S. DE et al. Significant differences in cesarean section rates between a private and a public hospital in Brazil. **Cadernos de Saúde Pública**, v. 24, n. 12, p. 2909–2918, dez. 2008.
- ANDREONI, M. **An Intrusion Detection And Prevention Architecture For Software Defined Networking**. [s.l: s.n.].
- BALAKRISHNAMA, S.; GANAPATHIRAJU, A. Linear discriminant analysis-a brief tutorial. **Institute for Signal and information Processing**, v. 18, p. 1–8, 1998.
- BARALDI, E. et al. Untargeted Metabolomic Analysis of Amniotic Fluid in the Prediction of Preterm Delivery and Bronchopulmonary Dysplasia. **PLOS ONE**, v. 11, n. 10, p. e0164211, 18 out. 2016.
- BARBOSA, D. C. M. **Fatores associados com altas taxas de cesáreas na coorte de nascimentos de Ribeirão Preto em 2010: projeto BRISA**. Ribeirão Preto: Universidade de São Paulo, 19 abr. 2017.
- BEDEEUZZAMAN, M.; ANJU, S. Preterm Birth Prediction Using EHG Signals. **International Journal of Scientific Research & Engineering Trends**, v. 5, n. 3, p. 3–6, 2019.
- BLENCOWE, H. et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. **The Lancet**, v. 379, n. 9832, p. 2162–2172, jun. 2012.
- BLENCOWE, H. et al. Born too soon: The global epidemiology of 15 million preterm births understanding the data preterm birth what is it? Defining preterm birth. **Reproductive Health**, v. 10, n. Suppl 1, p. 1–13, 2013a.
- BLENCOWE, H. et al. Preterm-associated visual impairment and estimates of retinopathy of prematurity at regional and global levels for 2010. **Pediatric Research**, v. 74, n. S1, p. 35–49, 20 dez. 2013b.
- BORG, D. **Redes neurais e support vector machines como técnicas de diagnósticos em medições industriais de nível por tecnologia tipo radar sem contato e apoio à decisão para a melhoria de sua aplicação**. São Carlos: Universidade de São Paulo, 22 fev. 2017.
- BRASIL. **PORTARIA Nº 570, DE 1º DE JUNHO DE 2000** Brasil, 2000. Disponível em: <[http://bvsmms.saude.gov.br/bvs/saudelegis/gm/2000/prt0570\\_01\\_06\\_2000\\_rep.html](http://bvsmms.saude.gov.br/bvs/saudelegis/gm/2000/prt0570_01_06_2000_rep.html)>
- BRASIL. **Datasus. Informações de Saúde**. Disponível em: <<http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6936>>. Acesso em: 7 dez. 2019.
- BREIMAN, L. **USING ADAPTIVE BAGGING TO DEBIAS REGRESSIONS**. [s.l: s.n.].
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5–32, 2001.
- BROWNE, M. W. Cross-Validation Methods. **Journal of Mathematical Psychology**, v. 44, n. 1, p. 108–132, mar. 2000.

- CAI, R. et al. Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. **Artificial Intelligence in Medicine**, v. 76, p. 7–15, fev. 2017.
- CARTER, J. et al. Development and validation of prediction models for the QUIPP App v.2: a tool for predicting preterm birth in women with symptoms of threatened preterm labor. **Ultrasound in Obstetrics & Gynecology**, 2019.
- CATLEY, C. et al. Predicting High-Risk Preterm Birth Using Artificial Neural Networks. **IEEE Transactions on Information Technology in Biomedicine**, v. 10, n. 3, p. 540–549, jul. 2006.
- CAVALCANTI, T. **Support Vector Machine**. Disponível em: <<https://www.codigofluente.com.br/aula-08-scikit-learn-maquina-de-vetores-de-suporte/>>. Acesso em: 20 dez. 2019.
- CHAPIESKI, M. L.; EVANKOVICH, K. D. Behavioral effects of prematurity. **Seminars in Perinatology**, v. 21, n. 3, p. 221–239, jun. 1997.
- CHEN, H.-Y. et al. Exploring the risk factors of preterm birth using data mining. **Expert Systems with Applications**, v. 38, n. 5, p. 5384–5387, maio 2011.
- CIABURRO, G. **Regression Analysis with R**. [s.l.] Packt Publishing, 2018.
- COPPER, R. L. et al. The preterm prediction study: Maternal stress is associated with spontaneous preterm birth at less than thirty-five weeks' gestation. **American Journal of Obstetrics and Gynecology**, v. 175, n. 5, p. 1286–1292, 1996.
- CREASY, R. K.; GUMMER, B. A.; LIGGINS, G. C. System for Predicting Spontaneous Preterm Birth. **Obstetrics & Gynecology**, v. 55, n. 6, 1980.
- CREVIER, D. **AI: the tumultuous history of the search for artificial intelligence**. [s.l.] Basic Books, 1993.
- DUDA, R.; STORK, D.; HART, P. **Pattern Classification**. 2. ed. [s.l.] Wiley-Interscience, 2000.
- DUDEK-SHRIBER, L. Parent Stress in the Neonatal Intensive Care Unit and the Influence of Parent and Infant Characteristics. **American Journal of Occupational Therapy**, v. 58, n. 5, p. 509–520, 1 set. 2004.
- EDWARDS, J. et al. Developmental Coordination Disorder in School-Aged Children Born Very Preterm and/or at Very Low Birth Weight: A Systematic Review. **Journal of Developmental & Behavioral Pediatrics**, v. 32, n. 9, p. 678–687, nov. 2011.
- ELITH, J.; LEATHWICK, J. R.; HASTIE, T. A working guide to boosted regression trees. **Journal of Animal Ecology**, v. 77, n. 4, p. 802–813, jul. 2008.
- ELOVITZ, M. A. et al. Can microRNA profiling in maternal blood identify women at risk for preterm birth? **American Journal of Obstetrics and Gynecology**, v. 212, n. 6, p. 782.e1-782.e5, jun. 2015.
- EVERITT, B. S. (1993). **Cluster Analysis**. Edward Arnold, 3a edição.
- FAROOQI, A. et al. Chronic Conditions, Functional Limitations, and Special Health Care Needs in 10- to 12-Year-Old Children Born at 23 to 25 Weeks' Gestation in the 1990s: A Swedish National Prospective Follow-up Study. **PEDIATRICS**, v. 118, n. 5, p. e1466–e1477, 1 nov. 2006.
- FERGUS, P. et al. Prediction of Preterm Deliveries from EHG Signals Using Machine Learning. **PLoS ONE**, v. 8, n. 10, p. e77154, 28 out. 2013.

- FERRERO, C. A. **Algoritmo kNN para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia**. São Carlos: Universidade de São Paulo, 4 mar. 2009.
- FORCADA-GUEX, M. et al. Prematurity, maternal posttraumatic stress and consequences on the mother-infant relationship. **Early Human Development**, v. 87, n. 1, p. 21–26, 2011.
- FRIEDMAN, J. H. Multivariate Adaptive Regression Splines. **The Annals of Statistics**, v. 19, n. 1, p. 1–67, mar. 1991.
- FRODI, A.; THOMPSON, R. Infants' Affective Responses in the Strange Situation: Effects of Prematurity and of Quality of Attachment. **Child Development**, v. 56, n. 5, p. 1280, out. 1985.
- GADI, M. F. A. **Uma comparação de métodos de classificação aplicados à detecção de fraude em cartões de crédito**. São Paulo: Universidade de São Paulo, 22 abr. 2008.
- GALLAGHER, C; MADDEN, M.G.; D'ARCY, B. **A Bayesian Classification Approach to Improving Performance for a Real-World Sales Forecasting Application**. In: 2015 IEEE 14th International Conference on Machine Learning and Applications, Florida, 2015.
- GOEDICKE-FRITZ, S. et al. Preterm Birth Affects the Risk of Developing Immune-Mediated Diseases. **Frontiers in Immunology**, v. 8, 9 out. 2017.
- GOLDBERG, S. et al. Maternal Behavior and Attachment in Low-Birth-Weight Twins and Singletons. **Child Development**, v. 57, n. 1, p. 34–46, fev. 1986.
- GOLDENBERG, R. L. et al. The Preterm Prediction Study: Fetal Fibronectin Testing and Spontaneous Preterm Birth. **Medicine**, v. 87, n. 5, p. 643–648, 1996.
- GOODWIN, L.; MAHER, S. **Data mining for preterm birth prediction**. Proceedings of the 2000 ACM symposium on Applied computing - SAC '00. **Anais...**New York, New York, USA: ACM Press, 2000Disponível em: <<http://portal.acm.org/citation.cfm?doid=335603.335680>>
- GROER, M. W. et al. Development of the preterm infant gut microbiome: a research priority. **Microbiome**, v. 2, n. 1, p. 38, 2014.
- GUNKO, V. O.; POGORELOVA, T. N.; LINDE, V. A. Proteomic Profiling of the Blood Serum for Prediction of Premature Delivery. **Bulletin of Experimental Biology and Medicine**, v. 161, n. 6, p. 829–832, 26 out. 2016.
- HADDERS-ALGRA, M. Two distinct forms of minor neurological dysfunction: perspectives emerging from a review of data of the Groningen Perinatal Project. **Developmental Medicine & Child Neurology**, v. 44, n. 08, 16 ago. 2002.
- HAMET, P.; TREMBLAY, J. Artificial intelligence in medicine. **Metabolism**, v. 69, p. S36–S40, abr. 2017.
- HEARST, M. A. et al. Support vector machines. **IEEE Intelligent Systems and their Applications**, v. 13, n. 4, p. 18–28, jul. 1998.
- HILL, J. L. et al. Prediction of preterm birth in symptomatic women using decision tree modeling for biomarkers. **American Journal of Obstetrics and Gynecology**, v. 198, n. 4, p. 468.e1–468.e9, abr. 2008.
- HOLDITCH-DAVIS, D. et al. Posttraumatic Stress Symptoms in Mothers of Premature Infants. **Journal of Obstetric, Gynecologic & Neonatal Nursing**, v. 32, n. 2, p. 161–171, mar. 2003.

HORNIK, C. P. et al. Early and late onset sepsis in very-low-birth-weight infants from a large group of neonatal intensive care units. **Early Human Development**, v. 88, p. S69–S74, maio 2012.

HOTHORN, T.; SEIBOLD, H.; ZEILEIS, A. **A Toolkit for Recursive Partytioning**CRAN, , 2019. Disponível em: <<https://cran.r-project.org/web/packages/partykit/partykit.pdf>>

HUTCHEON, J. A. et al. **Small size at birth or abnormal intrauterine growth trajectory: Which matters more for child growth?** *American Journal of Epidemiology*, v. 183, n. 12, p. 1107–1113, 2016.

IBGE. **Panorama - IBGE Cidades**. Disponível em: <<https://cidades.ibge.gov.br/brasil/sp/ribeirao-preto/panorama>>. Acesso em: 15 ago. 2019.

ICKOVICS, J. Group prenatal care and preterm birth weight: results from a matched cohort study at public clinics\*1. **Obstetrics & Gynecology**, v. 102, n. 5, p. 1051–1057, nov. 2003.

JACKSON, P. **Introduction to Expert Systems**. 3. ed. [s.l.] Addison-Wesley Longman Publishing Co., 1998.

JHA, A. K. et al. Use of Electronic Health Records in U.S. Hospitals. **New England Journal of Medicine**, v. 360, n. 16, p. 1628–1638, 16 abr. 2009.

KAPLAN, A.; HAENLEIN, M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. **Business Horizons**, v. 62, n. 1, p. 15–25, jan. 2019.

KELLER, J. M.; GRAY, M. R.; GIVENS, J. A. A fuzzy K-nearest neighbor algorithm. **IEEE Transactions on Systems, Man, and Cybernetics**, v. SMC-15, n. 4, p. 580–585, jul. 1985.

KLINE, R. Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence. **IEEE Annals of the History of Computing**, v. 33, p. 11, 2011.

KOULLALI, B. et al. Risk assessment and management to prevent preterm birth. **Seminars in Fetal and Neonatal Medicine**, v. 21, n. 2, p. 80–88, abr. 2016.

KUHN, M. et al. **Classification and Regression Training**CRAN, , 2020. Disponível em: <<https://cran.r-project.org/web/packages/caret/caret.pdf>>

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 27 maio 2015.

LEE, K. J. et al. The Clinical Usefulness of Predictive Models for Preterm Birth with Potential Benefits: A KOREan Preterm collaboratE Network (KOPEN) Registry-Linked Data-Based Cohort Study. **International journal of medical sciences**, v. 17, n. 1, p. 1–12, 2020.

LIAW, A.; WIENER, M. **Breiman and Cutler's Random Forests for Classification and Regression**, CRAN, 2018. Disponível em: <<https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>>

LINDSAY, R. K. et al. DENDRAL: A case study of the first expert system for scientific hypothesis formation. **Artificial Intelligence**, v. 61, n. 2, p. 209–261, jun. 1993.

LIOTTI, G. Trauma, dissociation, and disorganized attachment: Three strands of a single braid. **Psychotherapy: Theory, Research, Practice, Training**, v. 41, n. 4, p. 472–486, 2004.

LIU, L. et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. **The Lancet**, v. 388, n. 10063, p. 3027–3035, dez. 2016.

LOHR, S. IBM Is Counting on Its Bet on Watson, and Paying Big Money for It. **The New York Times**, 17 out. 2016.

MAGOULAS, R.; LORICA, B. **Introduction to Big Data**. 2. ed. [s.l.] O'Reilly Media, 2009.

MAIN, M.; HESSE, E. Parents' unresolved traumatic experiences are related to infant disorganized attachment status: Is frightened and/or frightening parental behavior the linking mechanism? In: **Attachment in the preschool years: Theory, research, and intervention**. The John D. and Catherine T. MacArthur Foundation series on mental health and development. Chicago, IL, US: University of Chicago Press, 1990. p. 161–182.

MATHEWSON, K. J. et al. Mental health of extremely low birth weight survivors: A systematic review and meta-analysis. **Psychological Bulletin**, v. 143, n. 4, p. 347–383, 2017.

MCCORDUCK, P. **Machines Who Think**. 2. ed. [s.l.] A K Peters/CRC Press, 2004.

MCDONALD, A. D. et al. Prematurity and work in pregnancy. **Occupational and Environmental Medicine**, v. 45, n. 1, p. 56–62, 1 jan. 1988.

MEDEIROS, M. N. L. et al. **Validity of pre and post-term birth rates based on the date of last menstrual period compared to early obstetric ultrasonography**. *Cadernos de saúde pública*, v. 31, n. 4, p. 885–90, 2015.

MEIS, P. J. et al. The preterm prediction study: Risk factors for indicated preterm births. **American Journal of Obstetrics and Gynecology**, v. 178, n. 3, p. 562–567, 1998.

MENON, R. et al. Multivariate adaptive regression splines analysis to predict biomarkers of spontaneous preterm birth. **Acta Obstetrica et Gynecologica Scandinavica**, v. 93, n. 4, p. 382–391, 2014.

MERCER, B. M. et al. The Preterm Prediction Study: Effect of gestational age and cause of preterm birth on subsequent obstetric outcome. **American Journal of Obstetrics and Gynecology**, v. 181, n. 5 I, p. 1216–1221, 1999.

METZ, C. Trump Signs Executive Order Promoting Artificial Intelligence. **The New York Times**, 2019.

METZ, J. **Interpretação de clusters gerados por algoritmos de clustering hierárquico**. Universidade de São Paulo, 2006.

MEYER, D. et al. **Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)** CRAN, 2019. Disponível em: <<https://cran.r-project.org/web/packages/e1071/e1071.pdf>>

MILES, M. S. et al. Perceptions of stress, worry, and support in Black and White mothers of hospitalized, medically fragile infants. **Journal of Pediatric Nursing**, v. 17, n. 2, p. 82–88, abr. 2002.

MILLER, R. A. Medical Diagnostic Decision Support Systems--Past, Present, And Future: A Threaded Bibliography and Brief Commentary. **Journal of the American Medical Informatics Association**, v. 1, n. 1, p. 8–27, 1 jan. 1994.

MORAES, L. M. V. DE et al. Fatores associados à peregrinação para o parto em São Luís (Maranhão) e Ribeirão Preto (São Paulo), Brasil: uma contribuição da coorte BRISA. **Cadernos de Saúde Pública**, v. 34, n. 11, 8 nov. 2018.

MORAVEC, H. The Great 1980s AI Bubble: A Review of The Brain Makers. **AI Magazine**, v. 15, n. 3 SE-Book Reviews, 15 set. 1994.

- MULLER-NIX, C. et al. Prematurity, maternal stress and mother–child interactions. **Early Human Development**, v. 79, n. 2, p. 145–158, set. 2004.
- MURTHY, S. K. Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey. **Data Mining and Knowledge Discovery**, v. 2, n. 4, p. 345–389, 1998.
- MWAMAKAMBA, L. W.; ZUCCHI, P. Cost estimate of hospital stays for premature newborns of adolescent mothers in a Brazilian public hospital. **Einstein (São Paulo)**, v. 12, n. 2, p. 223–229, 2014.
- NATIONAL RESEARCH COUNCIL. **Funding a revolution: government support for computing research**. [s.l.] Washington, D.C National Academy Press, 1999.
- NICULAESCU, O. **Classifying data with decision trees**. Disponível em: <<https://elf11.github.io/2018/07/01/python-decision-trees-acm.html>>. Acesso em: 28 jan. 2020.
- O’CONNOR, A. How Artificial Intelligence Could Transform Medicine. **The New York Times**, 2019.
- O’CONNOR, A. R. et al. Long-Term Ophthalmic Outcome of Low Birth Weight Children With and Without Retinopathy of Prematurity. **PEDIATRICS**, v. 109, n. 1, p. 12–18, 1 jan. 2002.
- OLIVEIRA, P. R. **Redes Neurais Artificiais para Extração de Características**. São Carlos: Universidade de São Paulo, 20 mar. 2018.
- OSCHINA. **K-Fold Cross Validation**. Disponível em: <<https://my.oschina.net/Bettyty/blog/751627>>. Acesso em: 27 jan. 2020.
- OSKOVI KAPLAN, Z. A.; OZGU-ERDINC, A. S. Prediction of Preterm Birth: Maternal Characteristics, Ultrasound Markers, and Biomarkers: An Updated Overview. **Journal of Pregnancy**, v. 2018, 2018.
- PAN, I. et al. Machine learning for social services: A study of prenatal case management in Illinois. **American Journal of Public Health**, v. 107, n. 6, p. 938–944, 2017.
- PAPIERNIK-BERKHAUER, E. Coefficient de risque d’accouchement prématuré (C.R.A.P.). **La Presse médicale (1893-1983)**, v. 77, n. 21, p. 793–794, 1969.
- PAPIERNIK, E.; KAMINSKI, M. Multifactorial study of the risk of prematurity at 32 weeks of gestation. I. A study of the frequency of 30 predictive characteristics. **Journal of Perinatal Medicine**, v. 2, n. 1, p. 30–36, jan. 1974.
- PAULA BARBOSA SOBRAL, A. **Previsão De Carga Horária Uma Nova Abordagem Por Árvore De Decisão**. Rio De Janeiro, Brazil: Pontifícia Universidade Católica Do Rio De Janeiro, 29 abr. 2003.
- PESAPANE, F. et al. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. **Insights into Imaging**, v. 9, n. 5, p. 745–753, 15 out. 2018.
- PETROU, S.; EDDAMA, O.; MANGHAM, L. A structured review of the recent literature on the economic consequences of preterm birth. **Archives of Disease in Childhood: Fetal and Neonatal Edition**, v. 96, n. 3, p. 225–232, 2011.
- PINTO, F. et al. Born Preterm: A Public Health Issue. **Portuguese Journal of Public Health**, v. 37, n. 1, p. 38–49, 2019.
- QUINLAN, J. R. Learning decision tree classifiers. **ACM Computing Surveys**, v. 28, n. 1, p. 71–72, 1 mar. 1996.

- RAJPURKAR, P. et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. 14 nov. 2017.
- RIPLEY, B. et al. **Support Functions and Datasets for Venables and Ripley's MASS** CRAN, 2019. Disponível em: <<https://cran.r-project.org/web/packages/MASS/MASS.pdf>>
- ROCHA, D. **Introdução a Machine Learning: Knn**. Disponível em: <[https://rpubs.com/davimat/knn\\_no\\_r](https://rpubs.com/davimat/knn_no_r)>. Acesso em: 25 jan. 2020.
- ROCHA, P. R. H. **Influência da prematuridade e restrição do crescimento intrauterino na habilidade motora na infância**. Ribeirão Preto: Universidade de São Paulo, 14 fev. 2019.
- RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach (2nd Edition)**. Second ed. [s.l: s.n.].
- SAFAVIAN, S. R.; LANDGREBE, D. A survey of decision tree classifier methodology. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 21, n. 3, p. 660–674, 1991.
- SAIGAL, S. et al. Physical Growth and Current Health Status of Infants Who Were of Extremely Low Birth Weight and Controls at Adolescence. **PEDIATRICS**, v. 108, n. 2, p. 407–415, 1 ago. 2001.
- SHANIBA ASMI, P.; SUBRAMANIAM, K.; IQBAL, N. V. Classification of Nonlinear Features of Uterine Electromyogram Signal Towards the Prediction of Preterm Birth. **IETE Journal of Research**, v. 0, n. 0, p. 1–10, 2019.
- SILVA, A. A. M. DA et al. Which factors could explain the low birth weight paradox? **Revista de Saúde Pública**, v. 40, n. 4, p. 648–655, ago. 2006.
- SINGER, L. T. Maternal Psychological Distress and Parenting Stress After the Birth of a Very Low-Birth-Weight Infant. **JAMA**, v. 281, n. 9, p. 799, 3 mar. 1999.
- SON, M.; MILLER, E. S. Predicting preterm birth: Cervical length and fetal fibronectin. **Seminars in Perinatology**, v. 41, n. 8, p. 445–451, 2017.
- TURING, A. M. I.—COMPUTING MACHINERY AND INTELLIGENCE. **Mind**, v. LIX, n. 236, p. 433–460, 1 out. 1950.
- VOVSHA, I. et al. Using Kernel Methods and Model Selection for Prediction of Preterm Birth. 27 jul. 2016.
- WATSON, H. A. et al. The QUIPP App: a safe alternative to a treat-all strategy for threatened preterm labor. **Ultrasound in Obstetrics and Gynecology**, v. 50, n. 3, p. 342–346, 2017.
- WHO. **Born Too Soon The Global Action Report on Preterm Birth**. [s.l: s.n.].
- WHO. **WHO recommendations on antenatal care for a positive pregnancy experience**. [s.l: s.n.].
- WHO. **Preterm birth**. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/preterm-birth>>. Acesso em: 18 mar. 2019.
- YOUNGE, N. et al. Survival and Neurodevelopmental Outcomes among Periviable Infants. **New England Journal of Medicine**, v. 376, n. 7, p. 617–628, 16 fev. 2017.
- YU, V. L. et al. Antimicrobial selection by a computer. A blinded evaluation by infectious diseases experts. **JAMA**, v. 242, n. 12, p. 1279–82, 21 set. 1979.
- ZHANG, H. The optimality of naive Bayes. **AA**, v. 1, n. 2, p. 3, 2004.





## APÊNDICE A – Tabelas complementares

### SVM

Predito (%)		
Real	0	1
0	1120 (73,2)	193 (12,6)
1	76 (4,9)	141 (9,2)
<b>Accuracy</b>	0,824	
<b>Sensitivity</b>	0,422	
<b>Specificity</b>	0,936	
<b>PPV</b>	0,649	
<b>NPV</b>	0,853	
<b>AUC</b>	0,818	

### LDA

Predito (%)		
Real	0	1
0	1288 (84,1)	25 (1,6)
1	177 (11,5)	40 (2,6)
<b>Accuracy</b>	0,868	
<b>Sensitivity</b>	0,615	
<b>Specificity</b>	0,879	
<b>PPV</b>	0,184	
<b>NPV</b>	0,980	
<b>AUC</b>	0,804	

### Random Forest

Predito (%)		
Real	0	1
0	1288 (84,1)	25 (1,6)
1	108 (7)	109 (7,1)
<b>Accuracy</b>	0,913	
<b>Sensitivity</b>	0,813	
<b>Specificity</b>	0,922	
<b>PPV</b>	0,502	
<b>NPV</b>	0,980	
<b>AUC</b>	0,842	

### KNN

Predito (%)		
Real	0	1
0	1299 (84,9)	148 (9,6)
1	37 (2,4)	46 (3)
<b>Accuracy</b>	0,879	
<b>Sensitivity</b>	0,237	
<b>Specificity</b>	0,972	
<b>PPV</b>	0,554	
<b>NPV</b>	0,897	
<b>AUC</b>	0,511	

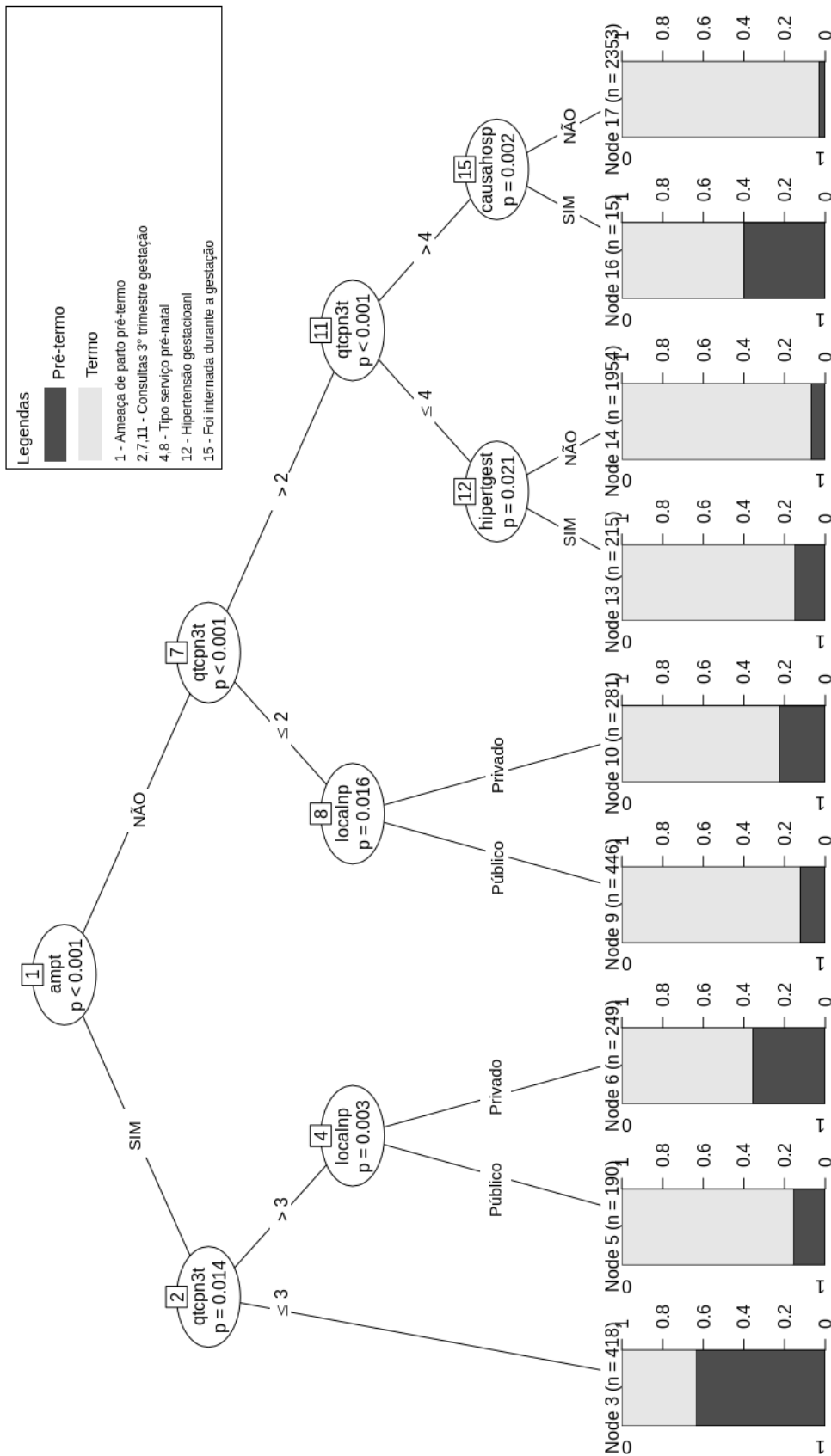
### Naive Bayes

Predito (%)		
Real	0	1
0	1293 (84,5)	20 (1,3)
1	191 (12,5)	26 (1,7)
<b>Accuracy</b>	0,862	
<b>Sensitivity</b>	0,565	
<b>Specificity</b>	0,871	
<b>PPV</b>	0,119	
<b>NPV</b>	0,984	
<b>AUC</b>	0,741	

### Decision Tree

Predito (%)		
Real	0	1
0	1246 (81,4)	100 (6,5)
1	83 (5,4)	101 (6,6)
<b>Accuracy</b>	0,880	
<b>Sensitivity</b>	0,502	
<b>Specificity</b>	0,937	
<b>PPV</b>	0,548	
<b>NPV</b>	0,925	
<b>AUC</b>	0,835	

APÊNDICE B – Árvore de decisão



## APÊNDICE C – Códigos (R codes)

### Análise de *Feature Selection* (MARS)

```
library(earth)

v2 <- earth(premacat ~ ., data=data_sc)

ev <- evimp(v2)

View(ev)
```

### Decision Tree

```
library(partykit)
library(plyr)
library(caret)

form <- "premacat ~
partoantpt+ampt+qtcpn3t+qtcpn2t+localnp+pn+malcoolg2t+causahosp+recanemia+doencagest
+hipertgest+nfilhos+idademae+habitofumo+paroutrab+fumograv+mes1cpn+nusprn"

set.seed(123)

folds <- split(data_na, cut(sample(1:nrow(data_na)),5))
errs.ctree <- rep(NA, length(folds))

for (i in 1:length(folds)) {
  test <- ldply(folds[i], data.frame)
  train <- ldply(folds[-i], data.frame)
  tmp.model <- ctree(as.formula(form), train, weights = ifelse(train$premacat=='0', 0.3, 1),
                    control = ctree_control(maxdepth = 4, multiway = TRUE))
  tmp.predict <- predict(tmp.model, newdata=test)
  conf.mat <- table(test$premacat, tmp.predict)
  errs.ctree[i] <- 1 - sum(diag(conf.mat))/sum(conf.mat)
}

print(sprintf("average error: %.3f percent", 100*mean(errs.ctree)))
plot(tmp.model)
confusionMatrix(conf.mat, positive = '1')
```

### Support Vector Machine

```
library(tidyverse)
library(caret)
library("e1071")

set.seed(123)

folds <- split(data, cut(sample(1:nrow(data)),5))
errs.svm <- rep(NA, length(folds))
wts <- 50 / table(train$premacat)

for (i in 1:length(folds)) {
  test <- ldply(folds[i], data.frame)
  train <- ldply(folds[-i], data.frame)
  tmp.model <- svm(premacat ~
partoantpt+ampt+qtcpn3t+qtcpn2t+localnp+pn+malcoolg2t+causahosp+recanemia+doencagest
+hipertgest+nfilhos+idademae+habitofumo+paroutrab+fumograv+mes1cpn+nusprn, train,
probability = TRUE, class.weights = wts, kernel = "radial", epsilon = 0.1, cost = 4)
```

```

tmp.predict <- predict(tmp.model, newdata=test)
conf.mat <- table(test$premacat, tmp.predict)
errs.svm[i] <- 1 - sum(diag(conf.mat))/sum(conf.mat)
}

tmodel <- tune(svm, premacat ~
partoantpt+ampt+qtcpn3t+fetos+localnp+recanemia+hipertgest+mothosp,
  data = train[, ], class.weights = wts, ranges = list(epsolon = seq(0,1,0.1), cost = 2^(1:6)))

plot(tmodel)
summary(tmodel)

model <- tmodel$best.model
tmp.predict <- predict(model, newdata=test)
conf.mat <- table(test$premacat, tmp.predict)

print(sprintf("average error: %.3f percent", 100*mean(errs.svm)))
confusionMatrix(conf.mat, positive = '1')

```

### Random Forest

```

library(randomForest)
library(plyr)

set.seed(123)

folds <- split(training, cut(sample(1:nrow(training)),5))
errs.rf <- rep(NA, length(folds))
wts <- 50 / table(training$premacat)

for (i in 1:length(folds)) {
  test <- ldply(folds[i], data.frame)
  train <- ldply(folds[-i], data.frame)
  tmp.model <- randomForest(premacat ~
partoantpt+ampt+qtcpn3t+qtcpn2t+localnp+pn+malcoolg2t+hosp+recanemia+outrad+hipertgest
+nfilhos+idademae+habitofumo+ativrem+fumograv+mes1cpn+nuspn,
  train, ntree=500, classwt = wts)
  tmp.predict <- predict(tmp.model, newdata=test)
  conf.mat <- table(test$premacat, tmp.predict)
  errs.rf[i] <- 1 - sum(diag(conf.mat))/sum(conf.mat)
}

print(sprintf("average error: %.3f percent", 100*mean(errs.rf)))
confusionMatrix(conf.mat, positive = '1')
plot(tmp.model)

```

### Naive Bayes

```

library("e1071")

set.seed(123)

folds <- split(data, cut(sample(1:nrow(data)),5))
errs.nb <- rep(NA, length(folds))
wts <- 50 / table(data$premacat)

for (i in 1:length(folds)) {
  test <- ldply(folds[i], data.frame)
  train <- ldply(folds[-i], data.frame)

```

```

tmp.model <- naiveBayes(premacat ~
partoantpt+ampt+qtcpn3t+qtcpn2t+localnp+pn+malcoolg2t+causahosp+recanemia+doencagest
+hipertgest+nfilhos+idademae+habitofumo+paroutrab+fumograv+mes1cpn+nuspn,
  train, weights = wts)
tmp.predict <- predict(tmp.model, newdata=test)
conf.mat <- table(test$premacat, tmp.predict)
errs.nb[i] <- 1 - sum(diag(conf.mat))/sum(conf.mat)
}

print(sprintf("average error: %.3f percent", 100*mean(errs.nb)))

```

```

confusionMatrix(conf.mat, positive = '1')

```

### Linear Discriminant Analysis

```

library(MASS)
library(plyr)

set.seed(123)

folds <- split(data, cut(sample(1:nrow(data)),5))
errs.lda <- rep(NA, length(folds))
wts <- 50 / table(data_na$premacat)

for (i in 1:length(folds)) {
  test <- ldply(folds[i], data.frame)
  train <- ldply(folds[-i], data.frame)
  tmp.model <- lda(premacat ~
partoantpt+ampt+qtcpn3t+qtcpn2t+localnp+pn+malcoolg2t+causahosp+recanemia+doencagest
+hipertgest+nfilhos+idademae+habitofumo+paroutrab+fumograv+mes1cpn+nuspn,
  data, method = "mle")
  tmp.predict <- predict(tmp.model, newdata=test)
  conf.mat <- table(test$premacat, tmp.predict$class)
  errs.lda[i] <- 1 - sum(diag(conf.mat))/sum(conf.mat)
}

print(sprintf("average error: %.3f percent", 100*mean(errs.lda)))
confusionMatrix(conf.mat, positive = '1')

```

### K-nearest neighbors

```

library(caret)

set.seed(123)

indxTrain <- createDataPartition(y = data$premacat, p = 0.8, list = FALSE)
training <- data[indxTrain,]
testing <- data[-indxTrain,]

prop.table(table(training$premacat)) * 100
prop.table(table(testing$premacat)) * 100

trainX <- training[,names(training) != "premacat"]

preProcValues <- preProcess(x = trainX,method = c("center", "scale"))

attach(training)

```

```
train =  
data.frame(partoantpt,ampt,qtcpn3t,qtcpn2t,localnp,pn,malcoolg2t,causahosp,reanemia,doenc  
agest,hipertgest,nfilhos,idademae,habitofumo,paroutrab,fumograv,mes1cpn,nuspn)  
  
training = data.frame(scale(train))  
  
model <- train(premacat ~ ., train, method = "knn", trControl = trainControl(method = "cv",  
number = 5, verboseIter = TRUE))  
  
knnPredict <- predict(model, newdata = testing)  
  
confusionMatrix(knnPredict, testing$premacat, positive = '1')
```

## APÊNDICE D – Modelo de predição categórico

O modelo de predição categórico foi desenvolvido com base, e a partir dos resultados do modelo de predição binário. O intuito do desenvolvimento deste modelo, é que em uma possível aplicação futura, o mesmo possa fornecer um resultado mais preciso do que somente uma resposta binária, se irá ocorrer um parto PT ou não.

Para tanto, foi utilizado no desenvolvimento do modelo categórico exatamente a mesma combinação de variáveis, resultante da análise de *feature selection* e revisão bibliográfica.

A variável desfecho foi criada com quatro categorias, termo ( $\geq 37$  semanas de IG), PT tardio (34 a 36 semanas e 6 dias de IG), PT moderado (31 a 33 semanas e 6 dias de IG), muito e extremos PT ( $< 31$  semanas de IG), unidos em uma única categoria devido ao baixo número de ocorrência. A frequência de cada categoria nessa classificação é de: 87,3% de partos termo, 9,5% PT tardio, 1,9% PT moderado, 1,3% muito e extremos PT.

O modelo foi desenvolvido em *Python*, dada a necessidade de integração com o protótipo da aplicação, o classificador escolhido foi *Support Vector Machine* (SVM), devido aos resultados observados e a modelagem seguiu a rigor todas as etapas e características que fora utilizado no modelo binário.

O modelo foi desenvolvido com uso de *5-folds cross validation*, balanceamento das classes (atribuindo pesos no treinamento, inversamente proporcionais à distribuição das classes) e ajuste de hiperparâmetros para obtenção do melhor resultado.

A tabela abaixo mostra a matriz de confusão resultante:

Real	Predito			
	0	1	2	3
0	1193	81	39	22
1	64	33	26	18
2	3	7	11	10
3	0	3	7	14

0 – Termo; 1 – PT tardio; 2 – PT moderado; 3 – Muito + extremo PT.

O modelo apresentou as seguintes características: 82% acurácia, sensibilidade 0,296, especificidade 0,893, valor positivo preditivo 0,22 e valor preditivo negativo 0,946. Visto que, os resultados obtidos pelos modelos binários são superiores ao encontrado aqui, e para que não haja discordância entre os resultados dos modelos de predição durante a utilização do aplicativo, uma regra lógica foi estabelecida.

Se o resultado da predição do modelo binário indicar um parto termo, automaticamente o modelo categórico não é acionado e assume-se como resposta um parto termo também. Caso o resultado da predição do modelo binário for PT, nessa situação o modelo categórico é executado e retorna o resultado da sua predição normalmente.



## APÊNDICE E – Protótipo da aplicação web (*backend*)

```

import numpy as np
import os
import joblib
from flask import Flask, request, render_template, make_response, url_for

app = Flask(__name__)
model = joblib.load('/home/gabriel/api/model/model.pkl')
model2 = joblib.load('/home/gabriel/api/model/model_c.pkl')

@app.route('/')
def display_gui():
    return render_template('home.html')
@app.route('/modelo', methods=['POST'])
def carrega_modelo():
    return render_template('template.html')
@app.route('/home', methods=['POST'])
def home():
    return render_template('home.html')
@app.route('/about', methods=['POST'])
def sobre():
    return render_template('about.html')
@app.route('/verificar', methods=['POST'])
def verificar():
    partoantptcat = request.form['partoantptcat']
    ampt = request.form['ampt']
    qtcpn3t = request.form['qtcpn3t']
    qtcpn2t = request.form['qtcpn2t']
    localnp = request.form['localnp']
    pn = request.form['pn']
    malcoolg2t = request.form['malcoolg2t']
    causahosp = request.form['causahosp']
    recanemia = request.form['recanemia']
    doencagest = request.form['doencagest']
    hipertgest = request.form['hipertgest']
    nfilhos = request.form['nfilhos']
    idademaie = request.form['idademaie']
    habitofumo = request.form['habitofumo']
    paroutrab = request.form['paroutrab']
    fumograv = request.form['fumograv']
    mes1cpn = request.form['mes1cpn']
    nuspn = request.form['nuspn']

    teste = np.array([[partoantptcat, ampt, qtcpn3t, qtcpn2t, localnp, pn, malcoolg2t,
causahosp, recanemia, doencagest, hipertgest, nfilhos, idademaie, habitofumo,
paroutrab, fumograv, mes1cpn, nuspn]])

    teste2 = np.array([[partoantptcat, ampt, qtcpn3t, qtcpn2t, localnp, pn,
causahosp, recanemia, doencagest, hipertgest, nfilhos, idademaie, habitofumo,
paroutrab, fumograv, mes1cpn, nuspn]])

    classe = model.predict(teste)[0]
    classe2 = model2.predict(teste2)[0]
    return render_template('pred_cl.html', classe =str(classe), classe2=str(classe2))

if __name__ == "__main__":
    port = int(os.environ.get('PORT', 5500))
    app.run(host='0.0.0.0', port=port, debug=True)

```

**Resultado:**

Tela inicial do protótipo (executado a partir do navegador, semelhante a um site):

## Modelo de aplicação Web para predição de Parto Pré-termo

*Web application model for predicting preterm birth*

---

Iniciar

Sobre

---

Apoio:



Tela Sobre (mais informações do estudo):

## Sobre o projeto:

*About this project:*

### Introdução:

O nascimento pré-termo (PT) é um dos principais fatores que levam a morte e causa, em longo prazo, significativa perda de capacidade, tanto física quanto psicológica entre os sobreviventes.

### Objetivos:

Desenvolver uma ferramenta computacional baseada em Aprendizado de Máquina capaz de prever o nascimento PT em uma coorte de nascimentos brasileira.

### Métodos:

Trata-se de um estudo transversal, baseado na coorte BRISA, que coletou informações sobre todos os nascimentos em 8 hospitais de Ribeirão Preto, Brasil em 2010. Foram aplicados questionários sobre a mãe e o recém-nascido. Nesta análise, apenas as informações maternas anteriores ao nascimento foram consideradas. Foram utilizadas características sociodemográficas, estilo de vida, características da gestação e acompanhamento pré-natal. Uma regressão não paramétrica foi utilizada para determinar as variáveis independentes e a variável dependente foi criada a partir da categorização da idade gestacional. Seis algoritmos de predição foram testados, Random Forest (RF), Support vector machines (SVM), Naive Bayes (NB), Decision Tree (DT), Linear discriminant analysis (LDA) e K-nearest neighbors (KNN).

[Voltar ao Início](#)

Apoio:



## Tela com formulário:

Idade da Mãe:	<input type="text" value="0"/>
N de Filhos:	<input type="text" value="0"/>
Parto ant PT	<input checked="" type="radio"/> Sim <input type="radio"/> Não
Ameaça parto PT	<input checked="" type="radio"/> Sim <input type="radio"/> Não
QTD consultas 2T:	<input type="text" value="0"/>
QTD consultas 3T:	<input type="text" value="0"/>
Média Alcool:	<input type="text" value="0"/>
Causa Hospitalização (Sem hosp = 0):	<input type="text" value="0"/>
Fez Pré-natal?	<input type="text" value="Sim"/>
Local PN	<input checked="" type="radio"/> Público <input type="radio"/> Privado <input type="radio"/> Não fez PN
Suplemento para Anemia:	<input type="radio"/> Sim <input checked="" type="radio"/> Não
Doença na gestação:	<input type="radio"/> Sim <input checked="" type="radio"/> Não
Hipertensão na gestação:	<input type="radio"/> Sim <input checked="" type="radio"/> Não
Fumou antes da gestação:	<input type="radio"/> Sim <input checked="" type="radio"/> Não
Fuma durante:	<input type="radio"/> Sim <input checked="" type="radio"/> Não
Semanas de Trabalho:	<input type="text" value="0"/>
Mês inicio PN	<input type="text" value="0"/>
N Ultrassons	<input type="text" value="0"/>

[Verificar](#)

Resultado da predição:

## Resultado da Predição

*Prediction Result:*

---

Modelo Binário:

**Pré-termo**

Modelo Categórico:

**34 a 36 semanas e 6 dias**

---

[Voltar](#)

[Página Inicial](#)


## ANEXO A

Parecer do Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (CEP-HCFMRP-USP) referente ao estudo - "Fatores Etiológicos do Nascimento e Consequências dos Fatores Perinatais na Saúde da Criança: Coorte de Nascimento".



HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA  
DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO

www.hcrp.fmrp.usp.br



Ribeirão Preto, 13 de novembro de 2008

Ofício nº 4116/2008  
CEP/RCC

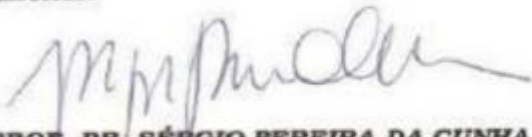
**Prezados Professores,**

O trabalho intitulado "**FATORES ETIOLÓGICOS DO NASCIMENTO PRÉ-TERMO E CONSEQUÊNCIAS DOS FATORES PERINATAIS NA SAÚDE DA CRIANÇA: COORTES DE NASCIMENTOS EM DUAS CIDADES BRASILEIRAS**", foi analisado pelo Comitê de Ética em Pesquisa, em sua 276ª Reunião Ordinária realizada em 10/11/2008, e enquadrado na categoria: **APROVADO, bem como o Termo de Consentimento Livre e Esclarecido.**

*Este Comitê segue integralmente a Conferência Internacional de Harmonização de Boas Práticas Clínicas (ICH-GCP), bem como a Resolução nº 196/96 CNS/MS.*

Lembramos que devem ser apresentados a este CEP, o Relatório Parcial e o Relatório Final da pesquisa.

Atenciosamente.



**PROF. DR. SÉRGIO PEREIRA DA CUNHA**  
Coordenador do Comitê de Ética em Pesquisa  
do HCRP e da FMRP-USP

---

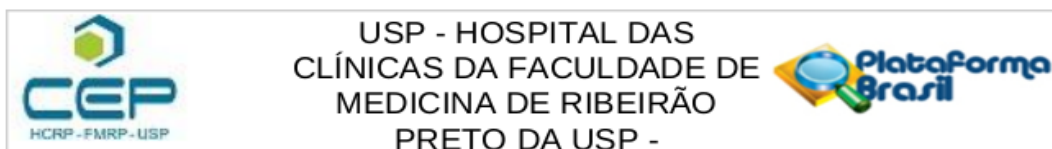
Ilustríssimos Senhores  
**PROF. DR. MARCO ANTONIO BARBIERI**  
**PROFª. DRª. HELOISA BETTIOL**  
Depto. de Puericultura e Pediatria

---

Comitê de Ética em Pesquisa HCRP e FMRP-USP- Campus Universitário.  
FWA-00002733; BRB-00002186 e Registro SISNEP/CONEP nº 4  
Fone (16) 3602-2228 - E-mail : cep@hcrp.fmrp.usp.br  
Monte Alegre 14048-900 Ribeirão Preto - SP

**ANEXO B**

Parecer do Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (CEP-HCFMRP-USP) referente ao presente estudo - “Desenvolvimento e análise de modelos de estatística multivariada baseados em Aprendizado de Máquina: predição de partos pré-termo na coorte de nascimento BRISA em Ribeirão Preto”.



Continuação do Parecer: 3.842.040

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1406712.pdf	10/02/2020 13:41:01		Aceito
Projeto Detalhado / Brochura Investigador	Projeto_mestrado_GabrielBazo.pdf	10/02/2020 13:40:05	GABRIEL BAZO	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Pedido_dispenza_TCLE_assinado.pdf	10/02/2020 13:39:31	GABRIEL BAZO	Aceito
Outros	AvaliaODaProposta_ComitDeRevis_Gabriel_Bazo.pdf	10/02/2020 09:57:32	GABRIEL BAZO	Aceito
Outros	CRC_DPP_Gabriel_Bazo.pdf	10/02/2020 09:56:40	GABRIEL BAZO	Aceito
Cronograma	Cronograma.pdf	10/02/2020 09:48:08	GABRIEL BAZO	Aceito
Outros	UPC_aprovado.pdf	10/02/2020 09:34:09	GABRIEL BAZO	Aceito
Orçamento	Orcamento_UPC.pdf	10/02/2020 09:31:59	GABRIEL BAZO	Aceito
Folha de Rosto	FR_10903.pdf	10/02/2020 09:17:17	GABRIEL BAZO	Aceito

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

RIBEIRAO PRETO, 17 de Fevereiro de 2020

Assinado por:  
**MARCIA GUIMARÃES VILLANOVA**  
(Coordenador(a))

Endereço: CAMPUS UNIVERSITÁRIO  
Bairro: MONTE ALEGRE CEP: 14.048-900  
UF: SP Município: RIBEIRAO PRETO  
Telefone: (16)3602-2228 Fax: (16)3633-1144 E-mail: cep@hcrp.usp.br

## ANEXO C – Questionário da Mãe



## QUESTIONÁRIO DO NASCIMENTO - MÃE

**BLOCO A – DADOS DE IDENTIFICAÇÃO****1A. Número de identificação:** \_\_\_\_\_

1ª casela: 1 Ribeirão Preto  
2 São Luís

2ª casela: 1 Pré-natal  
2 Nascimento  
3 1º ano

3ª casela: M. Avaliação no pré-natal  
A. Avaliação no nascimento RN 1  
B. Avaliação no nascimento RN 2  
C. Avaliação no nascimento RN 3  
D. Avaliação no nascimento RN 4

4ª e 5ª caselas: QM. Questionário da mãe  
QC. Questionário do RN  
SC. Saliva da criança  
CO. Cordão umbilical

6ª à 9ª. caselas: número seqüencial para cada cidade

NUMERO

         
**2A. Cidade:**

1.  Ribeirão Preto

2.  São Luís

CIDADE

**3A. Coorte**

1.  Iniciada no Pré-natal

2.  Iniciada no Nascimento

COORTE

**4A. Data da Entrevista (DD/MM/AAAA):** \_\_\_/\_\_\_/\_\_\_\_\_

DATAENT

Entrevistador (a) : \_\_\_\_\_

**5A. Hospital de Nascimento:**

SÃO LUÍS	RIBEIRÃO PRETO
1. <input type="checkbox"/> HU Materno- Infantil	12. <input type="checkbox"/> Hospital das Clínicas
2. <input type="checkbox"/> Benedito Leite	13. <input type="checkbox"/> Hospital Ribeirânia
3. <input type="checkbox"/> Marly Sarney	14. <input type="checkbox"/> Hospital São Lucas
4. <input type="checkbox"/> Santa Casa	15. <input type="checkbox"/> Hospital Santa Lydia
5. <input type="checkbox"/> Maria do Amparo	16. <input type="checkbox"/> Hospital Santa Casa
6. <input type="checkbox"/> N Sra. da Penha	17. <input type="checkbox"/> Mater
7. <input type="checkbox"/> Clínica São Marcos	18. <input type="checkbox"/> H. Sinhá Junqueira
8. <input type="checkbox"/> Clínica Luiza Coelho	19. <input type="checkbox"/> Hospital São Paulo
9. <input type="checkbox"/> Hospital S Domingos	
10. <input type="checkbox"/> Hospital Aliança	
11. <input type="checkbox"/> Clínica São José	

HOSPITAL



6A. Nome completo da mãe do RN (não abreviar):

\_\_\_\_\_  
\_\_\_\_\_

NOMEMAE

7A. Data de nascimento da mãe do RN (DD/MM/AAAA):

\_\_/\_\_/\_\_\_\_

DNMAE

8A. Idade da mãe do RN \_\_

99.  Não sabe

IDAEMA

9A. Qual a idade do pai do bebê? \_\_

99.  Não sabe

IDADEPAI

**BLOCO B – DADOS DE CONTATO**

1B. Qual o seu endereço completo? \_\_\_\_\_

\_\_\_\_\_

Telefone residencial: \_\_\_\_-\_\_\_\_ Outro telefone: \_\_\_\_-\_\_\_\_ celular: \_\_\_\_-\_\_\_\_

2B. Para facilitar futuros contatos, a sra. poderia nos fornecer o nome, relação de parentesco ou amizade, endereço e telefone fixo ou celular de parentes ou pessoas próximas com quem a sra. tem contato frequente?

Nome da pessoa: \_\_\_\_\_

Parentesco/Amizade: \_\_\_\_\_

Endereço: \_\_\_\_\_

\_\_\_\_\_

Telefone residencial: \_\_\_\_-\_\_\_\_ Telefone comercial: \_\_\_\_-\_\_\_\_ celular: \_\_\_\_-\_\_\_\_

Nome da pessoa: \_\_\_\_\_

Parentesco/Amizade: \_\_\_\_\_

Endereço: \_\_\_\_\_

\_\_\_\_\_

Telefone residencial: \_\_\_\_-\_\_\_\_ Telefone comercial: \_\_\_\_-\_\_\_\_ celular: \_\_\_\_-\_\_\_\_

3B. A sra. poderia nos fornecer o endereço e o telefone do seu trabalho?

Endereço: \_\_\_\_\_

\_\_\_\_\_

Telefone comercial: \_\_\_\_-\_\_\_\_ Telefone comercial: \_\_\_\_-\_\_\_\_

4B. Se a sra. pretende mudar de cidade, poderia nos informar o nome, endereço e o telefone de contato de algum parente ou alguém que more próximo à sua nova residência?

Nome da pessoa: \_\_\_\_\_

Endereço: \_\_\_\_\_

\_\_\_\_\_

Telefone residencial: \_\_\_\_-\_\_\_\_ Telefone comercial: \_\_\_\_-\_\_\_\_ celular: \_\_\_\_-\_\_\_\_

**BLOCO C – DADOS SOCIOECONÔMICOS E DEMOGRÁFICOS**

1C. A sra. sabe ler e escrever?

1.  Sim  
 2.  Não  
 9.  Não sabe

LERMAE

2C. A sra. frequenta ou frequentou escola?

1.  Sim  
 2.  Não **Passe para a questão 6C**  
 9.  Não sabe

ESCOLMAE

3C. A sra. ainda estuda ?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

MAEESTUDA

4C. Qual foi o último curso que a sra frequentou ou frequenta?

1.  Alfabetização de jovens e adultos  
 2.  Ensino fundamental ou 1o grau  
 3.  Ensino médio ou 2o grau  
 4.  Superior graduação incompleto **Passe para a questão 6C**  
 5.  Superior graduação completo **Passe para a questão 6C**  
 8.  Não se aplica  
 9.  Não sabe

CURSOMAE

5C. Até que série a sra. frequentou ou ainda frequenta?

1.  Primeira  
 2.  Segunda  
 3.  Terceira  
 4.  Quarta  
 5.  Quinta  
 6.  Sexta  
 7.  Sétima  
 8.  Oitava  
 88.  Não se aplica  
 99.  Não sabe

SERIEMAE

6C. Qual a cor da sua pele?

1.  branca  
 2.  preta/negra  
 3.  parda/mulata/cabocla/morena  
 4.  amarelo/oriental  
 5.  indígena  
 9.  não sabe

CORMAE

7C. Qual a situação conjugal atual da sra.?		
1. <input type="checkbox"/> Casada		
2. <input type="checkbox"/> União consensual (Mora junto)		
3. <input type="checkbox"/> Solteira		
4. <input type="checkbox"/> Separada/desquitada/divorciada		
5. <input type="checkbox"/> Viúva		
9. <input type="checkbox"/> Não sabe	SITCONMAE	<input type="checkbox"/>
8C. Quantas pessoas vivem atualmente na casa onde a sra. mora? (Considere apenas as pessoas que estão morando na casa há pelo menos 3 meses, e que não são temporários, como um tio que está temporariamente vivendo com a sra. por menos de 3 meses ou visitantes). __		
99. <input type="checkbox"/> Não sabe	PESSOAS	<input type="checkbox"/> <input type="checkbox"/>
9C. A sra. mora atualmente com o marido ou companheiro?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe	MORACOMP	<input type="checkbox"/>
10C. A sra. mora atualmente com filhos (biológicos ou não)?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passe para a questão 12C</b>		
9. <input type="checkbox"/> Não sabe	MORAFILHO	<input type="checkbox"/>
11C. Caso sim, com quantos filhos? __		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	QTFILHOS	<input type="checkbox"/> <input type="checkbox"/>
12C. A sra. mora atualmente com outros familiares?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passe para a questão 14C</b>		
9. <input type="checkbox"/> Não sabe	MORAFAM	<input type="checkbox"/>
13C. Caso sim, com quantos familiares? __		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	QTFAM	<input type="checkbox"/> <input type="checkbox"/>
14C. A sra. mora atualmente com outras pessoas que não são familiares?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passe para a questão 16C</b>		
9. <input type="checkbox"/> Não sabe	MORANFAM	<input type="checkbox"/>
15C. Caso sim, com quantos não familiares? __		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	QTNFAM	<input type="checkbox"/> <input type="checkbox"/>
16C. A sra. tem alguma religião ou culto?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe	TEMRELIG	<input type="checkbox"/>

17C. Caso tenha alguma religião, qual é a sua religião?

1.  Católica
2.  Evangélica. Ex: Batista, Assembléia de Deus, Bethesda, Universal, Adventista, Testemunha de Jeová, Luterana.
3.  Espírita/Kardecista
4.  Umbanda/Candomblé
5.  Judaica
6.  Orientais. Ex: Budista
7.  Outra, Qual? \_\_\_\_\_
88.  Não se aplica
99.  Não sabe

RELIGIAO

18C. A sra. exerce alguma atividade remunerada dentro ou fora de casa?

1.  Sim
2.  Não **Passa para a questão 26C**
9.  Não sabe

ATIVREM

**CASO NÃO TENHA ALGUMA ATIVIDADE REMUNERADA OU NÃO SAIBA, PASSE PARA A QUESTÃO 26C**

19C. Qual a sua ocupação ( o que faz atualmente no trabalho)?

- 
88.  Não se aplica
  99.  Não sabe

OCUPMAE

20C. Qual a sua relação de trabalho?

1.  Trabalha por conta própria
2.  Assalariado ou empregado
3.  Dono de empresa-empregador
4.  Faz bico
8.  Não se aplica
9.  Não sabe

RELACAO

21C. Quantos dias por semana a sra. trabalhava com remuneração durante a gestação? \_

8.  Não se aplica
9.  Não sabe

DIATRAB

22C. Nos dias de trabalho remunerado durante a gestação, quantas horas por dia a sra. trabalhava? \_\_\_

88.  Não se aplica
99.  Não sabe

HORATRAB

23C. Durante o seu trabalho, a sra. tinha que ficar em pé a maior parte do tempo?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

TRABPE

24C. Durante o seu trabalho, a sra. tinha que levantar coisas pesadas?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

LEVPESO

25C. Há quantas semanas atrás a sra. parou de trabalhar? \_\_

00.  menos de 1 semana  
 88.  Não se aplica  
 99.  Não sabe

PAROUTRAB

26C. Na sua casa, quem faz o trabalho de casa para sua família?

1.  A sra. faz todo trabalho  
 2.  A sra. faz parte do trabalho  
 3.  Outra pessoa  
 9.  Não sabe

TRABCASA

27C. Quem é a pessoa da família com maior renda atualmente? (considerar chefe da família aquele de maior renda)

1.  A entrevistada **Passa para a questão 36C**  
 2.  Companheiro  
 3.  Mãe  
 4.  Pai  
 5.  Avó  
 6.  Avô  
 7.  Madrasta  
 8.  Padrasto  
 9.  Tia  
 10.  Tio  
 11.  Irmã  
 12.  Irmão  
 13.  Outro \_\_\_\_\_  
 99.  Não sabe

CHEFE

28C. Qual o sexo da pessoa da família com maior renda?

1.  Masculino  
 2.  Feminino  
 8.  Não se aplica  
 9.  Não sabe

SEXOCHEFE

29C. Qual a idade da pessoa da família com maior renda (anos completos)? \_\_

88.  Não se aplica  
 99.  Não sabe

IDCHEFE

30C. Essa pessoa sabe ler e escrever?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

LERCHEFE

31C. Essa pessoa frequenta ou frequentou escola?

1.  Sim  
 2.  Não **Passe para a questão 34C**  
 8.  Não se aplica  
 9.  Não sabe

ESSCHEFE

32C. Qual foi o último curso que essa pessoa frequentou ou frequenta?

1.  Alfabetização de jovens e adultos  
 2.  Ensino fundamental ou 1o grau  
 3.  Ensino médio ou 2o grau  
 4.  Superior graduação incompleto **Passe para a questão 34C**  
 5.  Superior graduação completo **Passe para a questão 34C**  
 8.  Não se aplica  
 9.  Não sabe

CURSOCHefe

33C. Qual a série que essa pessoa frequenta ou até que série frequentou?

1.  Primeira  
 2.  Segunda  
 3.  Terceira  
 4.  Quarta  
 5.  Quinta  
 6.  Sexta  
 7.  Sétima  
 8.  Oitava  
 88.  Não se aplica  
 99.  Não sabe

SERIECHefe

34C. Qual a ocupação atual (ou no que trabalha) a pessoa com a maior renda da família? (Descreva a ocupação. Caso seja aposentado, colocar a última atividade que exerceu).

88.  Não se aplica

99.  Não sabe

OCUPCHEFE

35C. Qual a relação de trabalho do chefe da família?

1.  Trabalha por conta própria

2.  Assalariado ou empregado

3.  Dono de empresa-empregador

4.  Faz bico

8.  Não se aplica

9.  Não sabe

RELCHefe

36C. No mês passado quanto ganharam as pessoas da família que trabalham?

1ª pessoa R\$ \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_.

2ª pessoa R\$ \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_.

3ª pessoa R\$ \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_.

4ª pessoa R\$ \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_.

5ª pessoa R\$ \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_.

A família tem outra renda? \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_.

Renda total R\$ \_\_\_\_\_. \_\_\_\_\_. \_\_\_\_.

99999.  Não sabe

RENDAF

Quantos itens abaixo a família possui? (circule a resposta)

	Quantidade de itens				
	0	1	2	3	4 ou mais
37C. Televisão em cores	0	1	2	3	4
38C. Rádio	0	1	2	3	4
39C. Banheiro	0	4	5	6	7
40C. Automóvel	0	4	7	9	9
41C. Empregada mensalista	0	3	4	4	4
42C. Máquina de lavar	0	2	2	2	2
43C. Videocassete ou DVD	0	2	2	2	2
44C. Geladeira	0	4	4	4	4
45C. Freezer (aparelho independente ou parte da geladeira duplex)	0	2	2	2	2

TELEVISAO

RADIO

BANHEIRO

AUTOMOVEL

EMPREGADA

MAQLAVAR

DVD

GELADEIRA

FREEZER

**46C. Grau de Instrução da pessoa com maior renda**

Analfabeto/Primário incompleto/ Até 3ª Série Fundamental	0
Primário completo/ Até 4ª Série Fundamental/Ginasial incompleto	1
Ginasial completo/ Fundamental completo/Colegial incompleto	2
Colegial completo/ Médio completo/Superior incompleto	4
Superior completo	8

INSTRUCAO

**BLOCO D – HÁBITOS DE VIDA****Agora vamos conversar um pouco sobre o consumo de bebida alcoólica.**

1D. Durante a gravidez, a sra. tomou cerveja?

1.  Sim  
 2.  Não **Passe para a questão 14D**  
 9.  Não sabe

CERVEJA

2D. Durante a gravidez, a sra. tomou cerveja nos três primeiros meses de gravidez?

1.  Sim  
 2.  Não **Passe para a questão 6D**  
 8.  Não se aplica  
 9.  Não sabe

CERVEJA1T

3D. Quantos dias por semana? \_

8.  Não se aplica  
 9.  Não sabe

DIACERV1T

4D. Quanto tomava por dia (número de vasilhas) \_\_

88.  Não se aplica  
 99.  Não sabe

QTCERV1T

5D. Qual o tipo de vasilha?

1.  Copo comum (200ml)  
 2.  Lata (350ml)  
 3.  Garrafa pequena (300ml) – long neck  
 4.  Garrafa (600-720ml)  
 5.  Outro  
 8.  Não se aplica  
 9.  Não sabe

TIPOCERV1T



6D. Durante a gravidez a sra tomou cerveja dos 4 aos 6 meses de gravidez?

1.  Sim  
 2.  Não **Passe para a questão 10D**  
 8.  Não se aplica  
 9.  Não sabe

CERVEJA2T

7D. Quantos dias por semana? \_

8.  Não se aplica  
 9.  Não sabe

DIACERV2T

8D. Quanto tomava por dia (número de vasilhas) \_ \_

88.  Não se aplica  
 99.  Não sabe

QTCERV2T

9D. Qual o tipo de vasilha?

1.  Copo comum (200ml)  
 2.  Lata (350ml)  
 3.  Garrafa pequena (300ml) – long neck  
 4.  Garrafa (600-720ml)  
 5.  Outro  
 8.  Não se aplica  
 9.  Não sabe

TIPOCERV2T

10D. Durante a gravidez a sra. tomou cerveja dos 7 meses ao final da gravidez?

1.  Sim  
 2.  Não **Passe para a questão 14D**  
 8.  Não se aplica  
 9.  Não sabe

CERVEJA3T

11D. Quantos dias por semana? \_

8.  Não se aplica  
 9.  Não sabe

DIACERV3T

12D. Quanto tomava por dia (número de vasilhas) \_ \_

88.  Não se aplica  
 99.  Não sabe

QTCERV3T

13D. Qual o tipo de vasilha?

1.  Copo comum (200ml)  
 2.  Lata (350ml)  
 3.  Garrafa pequena (300ml) – long neck  
 4.  Garrafa (600-720ml)  
 5.  Outro  
 8.  Não se aplica  
 9.  Não sabe

TIPOCERV3T

14D. Durante a gravidez a sra tomou vinho?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passe para a questão 27D</b>		
9. <input type="checkbox"/> Não sabe	VINHO	<input type="checkbox"/>
15D. Durante a gravidez a sra tomou vinho nos três primeiros meses de gravidez?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passe para a questão 19D</b>		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	VINHO1T	<input type="checkbox"/>
16D. Quantos dias por semana? _		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	DIAVINHO1T	<input type="checkbox"/>
17D. Quanto tomava por dia (número de vasilhas) __		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	QTVINHO1T	<input type="checkbox"/> <input type="checkbox"/>
18D. Qual o tipo de vasilha?		
1. <input type="checkbox"/> Copo comum (200ml)		
2. <input type="checkbox"/> Cálice ou taça (400 ml)		
3. <input type="checkbox"/> Garrafa pequena (300ml)		
4. <input type="checkbox"/> Garrafa (600-720ml)		
5. <input type="checkbox"/> Outro		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	TIPOVINHO1T	<input type="checkbox"/>
19D. Durante a gravidez a sra. tomou vinho dos 4 aos 6 meses de gravidez?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passe para a questão 23D</b>		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	VINHO2T	<input type="checkbox"/>
20D. Quantos dias por semana? _		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	DIAVINHO2T	<input type="checkbox"/>
21D. Quanto tomava por dia (número de vasilhas) __		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	QTVINHO2T	<input type="checkbox"/> <input type="checkbox"/>

22D. Qual o tipo de vasilha?

1.  Copo comum (200ml)
2.  Cálice ou taça (400 ml)
3.  Garrafa pequena (300ml)
4.  Garrafa (600-720ml)
5.  Outro
8.  Não se aplica
9.  Não sabe

TIPOVINHO2T

23D. Durante a gravidez a sra. tomou vinho dos 7 meses ao final da gravidez?

1.  Sim
2.  Não **Passe para a questão 27D**
8.  Não se aplica
9.  Não sabe

VINHO3T

24D. Quantos dias por semana? \_

8.  Não se aplica
9.  Não sabe

DIAVINHO3T

25D. Quanto tomava por dia (número de vasilhas) \_ \_

88.  Não se aplica
99.  Não sabe

QTVINHO3T

26D. Qual o tipo de vasilha?

1.  Copo comum (200ml)
2.  Cálice ou taça (400 ml)
3.  Garrafa pequena (300ml)
4.  Garrafa (600-720ml)
5.  Outro
8.  Não se aplica
9.  Não sabe

TIPOVINHO3T

27D. Durante a gravidez a sra. tomou algum outro tipo de bebida como uisque, vodka, gim, rum?

1.  Sim
2.  Não **Passe para a questão 40D**
9.  Não sabe

DEST

28D. Durante a gravidez a sra. Tomou algum outro tipo de bebida como uisque, vodka, gim, rum nos primeiros meses de gravidez?

1.  Sim
2.  Não **Passe para a questão 32D**
8.  Não se aplica
9.  Não sabe

DEST1T

29D. Quantos dias por semana? \_

8.  Não se aplica
9.  Não sabe

DIADEST1T

30D. Quanto tomava por dia (número de vasilhas) \_\_

88.  Não se aplica

99.  Não sabe

QTDEST1T

31D. Qual o tipo de vasilha?

1.  Copo comum (200ml)

2.  Cálice, taça (400 ml)

3.  Martelo, copo de pinga (100ml)

4.  Lata (350ml) retirar

5.  Garrafa pequena (300ml)

6.  Garrafa (600-720ml)

7.  Outro

8.  Não se aplica

9.  Não sabe

TIPODEST1T

32D. Durante a gravidez a sra tomou algum outro tipo de bebida como uisque, vodka, gim, rum dos 4 aos 6 meses de gravidez?

1.  Sim

2.  Não **Passe para a questão 36D**

8.  Não se aplica

9.  Não sabe

DEST2T

33D. Quantos dias por semana? \_

8.  Não se aplica

9.  Não sabe

DIATEST2T

34D. Quanto tomava por dia (número de vasilhas) \_\_

88.  Não se aplica

99.  Não sabe

QTDEST2T

35D. Qual o tipo de vasilha?

1.  Copo comum (200ml)

2.  Cálice, taça (400 ml)

3.  Martelo, copo de pinga (100ml)

4.  Lata (350ml)

5.  Garrafa pequena (300ml)

6.  Garrafa (600-720ml)

7.  Outro

8.  Não se aplica

9.  Não sabe

TIPODEST2T

36D. Durante a gravidez a sra tomou algum outro tipo de bebida como uisque, vodka, gim, rum dos 7 meses ao final da gravidez?

1.  Sim

2.  Não **Passe para a questão 40D**

8.  Não se aplica

9.  Não sabe

DEST3T

37D. Quantos dias por semana? \_

8.  Não se aplica

9.  Não sabe

38D. Quanto tomava por dia (número de vasilhas) \_\_

88.  Não se aplica

99.  Não sabe

39D. Qual o tipo de vasilha?

1.  Copo comum (200ml)

2.  Cálice, taça (400 ml)

3.  Martelo, copo de pinga (100ml)

4.  Lata (350ml)

5.  Garrafa pequena (300ml)

6.  Garrafa (600-720ml)

7.  Outro

8.  Não se aplica

9.  Não sabe

DIADEST3T

QTDEST3T

TIPODEST3T

**Agora vamos conversar um pouco sobre o hábito de fumar.**

40D. A sra. tem ou teve o hábito de fumar cigarros?

1.  Sim

2.  Não **Passe para a questão 51D**

9.  Não sabe

41D. Com que idade a sra. começou a fumar cigarros? \_\_

88.  Não se aplica

99.  Não sabe

42D. Se a sra. parou de fumar, com que idade parou? \_\_

88.  Não se aplica ou ainda fuma

99.  Não sabe

43D. Se a sra. parou, quantos cigarros por dia em média a sra. costumava fumar? \_\_

88.  Não se aplica ou ainda fuma

99.  Não sabe

44D. No período de 6 meses antes desta gravidez a sra. fumava?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

45D. A sra. fumou durante esta gravidez?

1.  Sim

2.  Não **Passe para questão 51D**

8.  Não se aplica

9.  Não sabe

HABITOFUMO

IDAEFUMO

IDAEPAROU

NCIGPAROU

FUMOANTES

FUMOGRAV

46D. A sra. fumou do 1º ao 3º mês de gestação?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

FUMO1T

47D. A sra. fumou do 4º ao 6º mês de gestação?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

FUMO2T

48D. A sra. fumou do 7º mês de gestação até o final?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

FUMO3T

49D. Durante a gravidez a sra. fumava todos os dias?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

FUMODIA

50D. Quantos cigarros a sra. fumava por dia? \_\_\_

88.  Não se aplica

99.  Não sabe

QTFUMO

51D. A sra. convive em casa com outras pessoas que fumam?

1.  Sim

2.  Não **Passe para a questão 54D**

9.  Não sabe

FUMOCASA

52D. Quantas pessoas que residem com a sra. fumam? \_

8.  Não se aplica

9.  Não sabe

QTFUMCASA

53D. Estas pessoas fumam perto da sra. em sua casa?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

FUMOPERTO

54D. E no trabalho, as pessoas fumam perto da sra.?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

FUMOTRAB

55D. Caso a sra. tenha ficado próxima a pessoas que fumam, quantas horas por dia fica perto de fumantes?

1.  Menos de 1 hora por dia

HORASFUMO

2.  Mais de 1 hora por dia  
 8.  Não se aplica  
 9.  Não sabe

**Agora vamos conversar um pouco sobre o consumo de café.**

56D. A sra. costumava tomar café uma vez ou mais por semana durante a gravidez?

1.  Sim  
 2.  Não **Passe para a questão 1E**  
 9.  Não sabe

CAFE1T

57D. A sra. costumava tomar café uma vez ou mais por semana nos 3 primeiros meses da gravidez?

1.  Sim  
 2.  Não **Passe para a questão 61D**  
 8.  Não se aplica  
 9.  Não sabe

CAFE1T

58D. Quantos dias por semana a sra. tomava café? \_

8.  Não se aplica  
 9.  Não sabe

DIASCAFE1T

59D. Quantas vezes por dia a sra. tomava café? \_ \_

88.  Não se aplica  
 99.  Não sabe

VEZESCAFE1T

60D. Qual o tipo de vasilha em que a sra. costumava tomar café?

1.  xícara de chá  
 2.  xícara de cafezinho  
 3.  meia taça  
 4.  copo comum – 200 ml  
 5.  outro \_\_\_\_\_  
 8.  Não se aplica  
 9.  Não sabe

QTCAFE1T

61D. A sra. costumava tomar café uma vez ou mais por semana dos 4 aos 6 meses da gravidez?

1.  Sim  
 2.  Não **Passe para a questão 65D**  
 8.  Não se aplica  
 9.  Não sabe

CAFE2T

62D. Quantos dias por semana a sra. tomava café? \_

8.  Não se aplica  
 9.  Não sabe

DIASCAFE2T

63D. Quantas vezes por dia a sra. tomava café? \_ \_

88.  Não se aplica  
 99.  Não sabe

VEZESCAFE2T

64D. Qual o tipo de vasilha em que a sra. costumava tomar café?

1.  xícara de chá
2.  xícara de cafezinho
3.  meia taça
4.  copo comum - 200 ml
5.  outro \_\_\_\_\_
8.  Não se aplica
9.  Não sabe

QTCAFE2T

65D. A sra. costumava tomar café uma vez ou mais por semana dos 7 meses ao final da gravidez?

1.  Sim
2.  Não **Passa para a questão 1E**
8.  Não se aplica
9.  Não sabe

CAFE3T

66D. Quantos dias por semana a sra. tomava café? \_

8.  Não se aplica
9.  Não sabe

DIASCAFE3T

67D. Quantas vezes por dia a sra. tomava café? \_\_

88.  Não se aplica
99.  Não sabe

VEZESCAFE3T

68D. Qual o tipo de vasilha em que a sra. costumava tomar café?

1.  xícara de chá
2.  xícara de cafezinho
3.  meia taça
4.  copo comum - 200 ml
5.  outro
8.  Não se aplica
9.  Não sabe

QTCAFE3T

### **BLOCO E – DADOS DO COMPANHEIRO**

Caso o companheiro seja a pessoa com maior renda ou não more junto com a mãe do RN, passe para a questão 1F.

1E. Qual a idade do companheiro atual? \_\_

88.  Não se aplica – não tem companheiro atual
99.  Não sabe

IDCOMP

2E. O seu companheiro sabe ler e escrever?

1.  Sim
2.  Não
8.  Não se aplica
9.  Não sabe

LERCOMP



3E. O seu companheiro frequenta ou frequentou escola?

1.  Sim

2.  Não **Passe para a questão 6E**

8.  Não se aplica

9.  Não sabe

ESSCOMP

4E. Qual o último curso que seu companheiro frequentou ou frequenta?

1.  Alfabetização de jovens e adultos

2.  Ensino fundamental ou 1o grau

3.  Ensino médio ou 2o grau

4.  Superior graduação incompleto **Passe para a questão 6E**

5.  Superior graduação completo **Passe para a questão 6E**

8.  Não se aplica

9.  Não sabe

CURSOCOMP

5E. Qual a série que seu companheiro frequenta ou até que série estudou?

1.  Primeira

2.  Segunda

3.  Terceira

4.  Quarta

5.  Quinta

6.  Sexta

7.  Sétima

8.  Oitava

88.  Não se aplica

99.  Não sabe

SERIECOMP

6E. O seu companheiro está trabalhando no momento?

1.  Sim

2.  Não **Passe para a questão 1F**

8.  Não se aplica

9.  Não sabe

TRABCOMP

7E. Qual a ocupação atual (ou no que trabalha) o seu companheiro atual? (Descreva a ocupação. Caso seja aposentado, colocar a última atividade que exerceu).

88.  Não se aplica

99.  Não sabe

OCUPCOMP

8E. Qual a relação de trabalho do seu companheiro?

1.  Trabalha por conta própria

2.  Assalariado ou empregado

3.  Dono de empresa-empregador

4.  Faz bico

8.  Não se aplica

9.  Não sabe

RELCOMP

**BLOCO F – DADOS DA SAÚDE SEXUAL E REPRODUTIVA**

1F. Em que mês de gestação a sra. soube que estava grávida? _		
9. <input type="checkbox"/> Não sabe	MESGRAV	<input type="checkbox"/>
2F. Que idade a sra. tinha quando menstruou pela primeira vez? __		<input type="checkbox"/> <input type="checkbox"/>
99. <input type="checkbox"/> Não sabe	IDMENARCA	<input type="checkbox"/> <input type="checkbox"/>
3F. Qual era sua idade quando a sra. teve sua primeira relação sexual? __		<input type="checkbox"/> <input type="checkbox"/>
99. <input type="checkbox"/> Não sabe	IDRELSEX	<input type="checkbox"/> <input type="checkbox"/>
4F. Que idade a sra. tinha quando engravidou pela primeira vez? __		<input type="checkbox"/> <input type="checkbox"/>
99. <input type="checkbox"/> Não sabe	IDGRAV	<input type="checkbox"/> <input type="checkbox"/>
5F. Quantas vezes a sra. engravidou? (incluindo a gravidez atual) Contar todas as gestações até as que não chegaram no final, inclusive abortos. __		<input type="checkbox"/> <input type="checkbox"/>
99. <input type="checkbox"/> Não sabe	GESTA	<input type="checkbox"/> <input type="checkbox"/>
6F. A gravidez atual foi planejada?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe	PLANGEST	<input type="checkbox"/>
7F. Quantos filhos nasceram vivos (incluindo o atual)? __		<input type="checkbox"/> <input type="checkbox"/>
99. <input type="checkbox"/> Não sabe	NASCVIVO	<input type="checkbox"/> <input type="checkbox"/>
8F. Qual a idade da sra. no início da gestação anterior? __		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	IDULTGEST	<input type="checkbox"/> <input type="checkbox"/>
9F. Quantos partos a sra. teve (incluindo o atual)? __		<input type="checkbox"/> <input type="checkbox"/>
99. <input type="checkbox"/> Não sabe	PARTOS	<input type="checkbox"/> <input type="checkbox"/>
10F. Quantos partos foram cesáreas (incluindo o atual, caso tenha sido cesárea)? _		<input type="checkbox"/>
9. <input type="checkbox"/> Não sabe	PARTOCESA	<input type="checkbox"/>
11F. O parto anterior da sra. foi:		
1. <input type="checkbox"/> Vaginal/Vaginal com fórceps		
2. <input type="checkbox"/> Cesárea		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	TIPARTOANT	<input type="checkbox"/>
12F. Quantos partos anteriores da sra. foram antes do tempo (prematuros) incluindo o atual? _		<input type="checkbox"/>
9. <input type="checkbox"/> não sabe	PARTOANTPT	<input type="checkbox"/>
13F. A sra. teve algum aborto?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passe para questão 15F</b>		
9. <input type="checkbox"/> Não sabe	ABORTO	<input type="checkbox"/>
14F. Caso sim, quantos abortos foram? _		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	QTABORTO	<input type="checkbox"/>

15F. A sra teve algum filho que nasceu morto (incluir atual)?

1.  Sim

2.  Não **Passa para a questão 17F**

9.  Não sabe

NASCMORTO

16F. Caso sim, quantos foram (incluir atual)? \_

8.  Não se aplica

9.  Não sabe

QTMORTO

17F. Quantos filhos vivos a sra. tem atualmente? \_\_

99.  Não sabe

NFILHOS



### **BLOCO G - MORBIDADES**

**Agora vamos conversar com a sra. sobre os problemas que teve durante a gravidez.**

1G. A sra. teve hipertensão (pressão alta) fora da gestação diagnosticada por médico ou enfermeiro?

1.  Sim

2.  Não

9.  Não sabe

HIPERT

2G. A sra. teve hipertensão (pressão alta) na gestação atual?

1.  Sim

2.  Não

9.  Não sabe

HIPERTGEST

3G. A sra. teve antes da gestação nível elevado de açúcar no sangue (diabetes) diagnosticado por médico ou enfermeiro?

1.  Sim

2.  Não

9.  Não sabe

DIABETES

4G. A sra. teve nível elevado de açúcar no sangue (diabetes) diagnosticado por médico ou enfermeiro durante a gestação?

1.  Sim

2.  Não

9.  Não sabe

DIABGEST

5G. A sra. teve herpes durante a gestação diagnosticado por médico, enfermeiro ou dentista?

1.  Sim

2.  Não

9.  Não sabe

HERPESGEST

6G. A sra. teve sarampo durante a gestação diagnosticado por médico, enfermeiro ou dentista?

1.  Sim

2.  Não

9.  Não sabe

SARAMPGEST

7G. A sra. teve catapora durante a gestação diagnosticado por médico, enfermeiro ou dentista?

1.  Sim  
 2.  Não  
 9.  Não sabe

CATAPGEST

8G. A sra. teve rubéola durante a gestação diagnosticado por médico, enfermeiro ou dentista?

1.  Sim  
 2.  Não  
 9.  Não sabe

RUBGEST

9G. A sra. teve algum episódio de febre alta (temperatura acima de 38°) que durou mais de 24 horas durante esta gestação diagnosticada por médico, enfermeiro ou dentista?

1.  Sim  
 2.  Não **Passe para a questão 11G**  
 9.  Não sabe

FEBREGEST

10G. Caso sim, quantas vezes? \_\_

88.  Não se aplica  
 99.  Não sabe

QTFEBRE

11G. A sra. teve anemia antes da gestação diagnosticado por médico ou enfermeiro?

1.  Sim  
 2.  Não  
 9.  Não sabe

ANEMIA

12G. A sra. teve anemia diagnosticado por médico ou enfermeiro durante a gestação?

1.  Sim  
 2.  Não  
 9.  Não sabe

ANEMIAGEST

13G. A sra. teve toxoplasmose antes da gestação diagnosticado por médico ou enfermeiro?

1.  Sim  
 2.  Não  
 9.  Não sabe

TOXO

14G. A sra. teve toxoplasmose diagnosticado por médico ou enfermeiro durante a gestação?

1.  Sim  
 2.  Não  
 9.  Não sabe

TOXOGEST

15G. A sra. teve sífilis antes da gestação diagnosticado por médico ou enfermeiro?

1.  Sim  
 2.  Não  
 9.  Não sabe

SIFILIS

16G. A sra. teve sífilis durante a gestação diagnosticado por médico ou enfermeiro?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		
	<b>SIFILISGEST</b>	<input type="checkbox"/>
17G. A sra. teve alguma infecção urinária/ durante a gestação atual diagnosticada por médico ou enfermeiro?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		
	<b>INFECURI</b>	<input type="checkbox"/>
18G. A sra. teve algum corrimento vaginal durante a gestação atual?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		
	<b>CORRIMENTO</b>	<input type="checkbox"/>
19G. A sra. sofreu alguma queda ou acidente durante a gestação?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		
	<b>QUEDA</b>	<input type="checkbox"/>
20G. A sra. teve algum sangramento vaginal nos últimos 3 meses da gestação atual?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		
	<b>SANGVAGINA</b>	<input type="checkbox"/>
21G. A sra. foi internada alguma vez durante a gestação atual seja por qualquer motivo?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passe para a questão 23G</b>		
9. <input type="checkbox"/> Não sabe		
	<b>HOSP</b>	<input type="checkbox"/>
22G. Qual foi o motivo da internação?		
<hr/>		
88. <input type="checkbox"/> Não se aplica		
98. <input type="checkbox"/> Não sabe	<b>CAUSAHOSP</b>	<input type="checkbox"/> <input type="checkbox"/>
23G. A sra. teve ameaça de aborto na gestação atual?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		
	<b>AMABORTO</b>	<input type="checkbox"/>
24G. A sra. teve ameaça de parto prematuro (antes do tempo) na gestação atual?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		
	<b>AMPT</b>	<input type="checkbox"/>

25G. A sra. teve outra doença durante a gestação atual?

1.  Sim

2.  Não **Passe para a questão 27G**

9.  Não sabe

26G. Qual doença? \_\_\_\_\_

88.  Não se aplica

99.  Não sabe

OUTRAD

DOENCAGEST

27G. A sra. usou algum remédio durante a gravidez?

1.  Sim

2.  Não **Passe para o bloco H**

9.  Não sabe

REMGEST

Caso tenha utilizando, qual o nome do remédio e qual o mês da gestação a sra. estava quando começou e quando parou (mês da gestação)? Se continua usando, anotar que está em uso?

28G. Remédio

88.  Não se aplica

99.  Não sabe

NOMEREM1

29G. Mês de início\_

0.  Usa desde antes da gravidez

88.  Não se aplica

99.  Não sabe

MESIREM1

30G. Mês de término \_

10.  Ainda usa

88.  Não se aplica

99.  Não sabe

MESTREM1

31G. Remédio

88.  Não se aplica

99.  Não sabe

NOMEREM2

32G. Mês de Início \_

0.  Usa desde antes da gravidez

88.  Não se aplica

99.  Não sabe

MESIREM2

33G. Mês de término \_

10.  Ainda usa

88.  Não se aplica

9.9  Não sabe

MESTREM2

## 34G. Remédio

---

88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe		
35G. Mês de início _		
0. <input type="checkbox"/> Usa desde antes da gravidez		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	<b>NOMEREM3</b>	<input type="checkbox"/> <input type="checkbox"/>
36G. Mês de término _		
10. <input type="checkbox"/> Ainda usa		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	<b>MESIREM3</b>	<input type="checkbox"/> <input type="checkbox"/>
	<b>MESTREM3</b>	<input type="checkbox"/> <input type="checkbox"/>
37G. Remédio		

---

88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe		
38G. Mês de início _		
0. <input type="checkbox"/> Usa desde antes da gravidez		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	<b>NOMEREM4</b>	<input type="checkbox"/> <input type="checkbox"/>
39G. Mês de término _		
10. <input type="checkbox"/> Ainda usa		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	<b>MESIREM4</b>	<input type="checkbox"/> <input type="checkbox"/>
	<b>MESTREM4</b>	<input type="checkbox"/> <input type="checkbox"/>
40G. Remédio		

---

88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe		
41G. Mês de início _		
0. <input type="checkbox"/> Usa desde antes da gravidez		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	<b>NOMEREM5</b>	<input type="checkbox"/> <input type="checkbox"/>
42G. Mês de término _		
10. <input type="checkbox"/> Ainda usa		
88. <input type="checkbox"/> Não se aplica		
99. <input type="checkbox"/> Não sabe	<b>MESIREM5</b>	<input type="checkbox"/> <input type="checkbox"/>
	<b>MESTREM5</b>	<input type="checkbox"/> <input type="checkbox"/>

**BLOCO H – CARACTERÍSTICAS DA GESTAÇÃO ATUAL E DO PRÉ-NATAL**

Pergunte se a puérpera dispõe do cartão da gestante e se está de posse do mesmo. Confirme as respostas no cartão

1H. A sra tem cartão da gestante?

1.  Sim  
 2.  Não  
 8.  Não se aplica – não fez pré-natal  
 9.  Não sabe

CARTAO

2H. Qual a data da sua última menstruação (DD/MM/AA)?

\_\_/\_\_/\_\_\_\_

99999999.  Não sabe

DUM

3H. Caso não saiba a data, informar o mês/ano: \_\_/\_\_\_\_

888888.  Não se aplica

999999.  Não sabe

MESANOUM

4H. Qual o seu peso antes de engravidar? \_\_\_\_\_, \_ kg

9999.  Não sabe

PESOANTES

5H. Qual a sua altura antes de engravidar? \_\_\_\_\_, \_ cm

9999.  Não sabe

ALTURAANT

6H. A sra fez pré-natal?

1.  Sim  
 2.  Não **Passe para questão 39H**  
 9.  Não sabe

PN

7H. Qual a data da primeira consulta pré-natal (DD/MM/AA)?

\_\_/\_\_/\_\_\_\_

88888888.  Não se aplica

99999999.  Não sabe

DT1CPN

8H. Em que mês de gravidez a sra. iniciou as consultas de pré-natal? \_\_

88.  Não se aplica

99.  Não sabe

MES1CPN

9H. Quantas consultas de pré-natal a sra. fez no 1º trimestre de a gestação? \_\_

88.  Não se aplica

99.  Não sabe

QTCPN1T

10H. Quantas consultas de pré-natal a sra. fez no 2º trimestre de a gestação? \_\_

88.  Não se aplica

99.  Não sabe

QTCPN2T

11H. Quantas consultas de pré-natal a sra. fez no 3º trimestre de a gestação? \_\_

88.  Não se aplica

99.  Não sabe

QTCPN3T



12H. O médico ou enfermeiro encaminhou a senhora para fazer tratamento com o dentista durante esta gravidez?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

TRATDEN

13H. Quantas consultas médicas a sra. fez no pré-natal? \_\_

88.  Não se aplica  
 99.  Não sabe

QTCPNMED

14H. Quantas consultas com enfermeira(o) a sra. fez no pré-natal? \_\_

88.  Não se aplica  
 99.  Não sabe

QTCPNENF

15H. Quantos exames de radiografia (incluindo radiografias dos dentes) foram feitos durante a gestação atual? \_\_

88.  Não se aplica  
 99.  Não sabe

QTEXRADI

16H. Em que local a sra. fez o pré-natal?

1.  SUS  
 3.  Plano de saúde/ seguro saúde  
 4.  Particular  
 8.  Não se aplica  
 9.  Não sabe

LOCALNP

17H. Qual a data da última consulta pré-natal (DD/MM/AA)?

\_\_/\_\_/\_\_\_\_

88888888.  Não se aplica  
 99999999.  Não sabe

DTUCPN

18H. Caso não saiba a data, informar o mês de gravidez aproximado: \_

8.  Não se aplica  
 9.  Não sabe

MESUCPN

19H. Quantos exames de ultrassonografia foram feitos durante a gestação atual? \_

8.  Não se aplica  
 9.  Não sabe

NUSPN

**Durante as consultas de pré-natal o médico ou enfermeira alguma vez:**

20H. Solicitou exame de sangue?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

EXSANGUE

21H. Solicitou exame de urina?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

**EXURINA**

22H. Perguntou a data da última menstruação?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

**PDUM**

23H. Verificou o seu peso?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

**EXPESO**

24H. Mediu a sua barriga?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

**EXBARRIGA**

25H. Receitou Cálcio?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

**RECCALCIO**

26H. Mediu a sua pressão?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

**EXPA**

27H. Fez exame ginecológico?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

**EXGINEC**

28H. Receitou remédio para anemia?

1.  Sim  
 2.  Não  
 8.  Não se aplica  
 9.  Não sabe

**RECANEMIA**

29H. Receitou vitamina?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	<b>RECVIT</b>	<input type="checkbox"/>
30H. Orientou sobre amamentação?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	<b>OAMAMENT</b>	<input type="checkbox"/>
31H. O médico perguntou se a sra. estava usando algum medicamento?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	<b>OMEDIC</b>	<input type="checkbox"/>
32H. Orientou sobre o risco do uso de remédios sem orientação médica durante a gravidez?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	<b>OREM</b>	<input type="checkbox"/>
33H. Orientou sobre como evitar toxoplasmose durante a gravidez? (lavar muito bem frutas e verduras, não comer carne mal passada, evitar contato com gatos, não manipular terra, lavar muito bem as mãos antes das refeições).		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	<b>OTOXO</b>	<input type="checkbox"/>
34H. Examinou o seu seio?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	<b>EXSEIO</b>	<input type="checkbox"/>
35H. Fez exame de prevenção de câncer de colo de útero?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	<b>EXCOLOUT</b>	<input type="checkbox"/>
36H. Fez exame de sangue para sífilis no pré-natal?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe	<b>EXSIFILIS</b>	<input type="checkbox"/>

37H. Fez exame de sangue para saber o tipo de sangue?		
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe		<input type="checkbox"/>
38H. Ofereceu exame de sangue para HIV no pré-natal?	<b>EXTIPOSANG</b>	
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe		<input type="checkbox"/>
39H. Você tomou vacina contra o tétano?	<b>EXHIV</b>	
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passa para a questão 41H</b>		
3. <input type="checkbox"/> Já estava vacinada antes da gravidez. <b>Passa para a questão 41H</b>		
9. <input type="checkbox"/> Não sabe		<input type="checkbox"/>
40H. Quantas doses de antitetânica a sra. recebeu? _	<b>VACTET</b>	
8. <input type="checkbox"/> Não se aplica		
9. <input type="checkbox"/> Não sabe		<input type="checkbox"/>
41H. Durante a gestação atual, a sra. recebeu atendimento de auxiliar de enfermagem?	<b>QTDOSSETET</b>	
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		<input type="checkbox"/>
42H. Durante a gestação atual, a sra. recebeu visita do agente de saúde?	<b>AUXENFPN</b>	
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		<input type="checkbox"/>
43H. Durante a gestação atual, a sra. recebeu atendimento de parteira leiga?	<b>AGSAUDEPN</b>	
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		<input type="checkbox"/>
44H. Durante a gestação atual, a sra. recebeu atendimento do programa de Saúde da Família (PSF)?	<b>PARTPN</b>	
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não		
9. <input type="checkbox"/> Não sabe		<input type="checkbox"/>
45H. Qual o seu peso ao final da gravidez? ____,_ kg	<b>PSFPN</b>	
9999. <input type="checkbox"/> Não sabe		<input type="checkbox"/>
46H. A sra. fez algum tratamento para engravidar?	<b>PESOFINAL</b>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
1. <input type="checkbox"/> Sim		
2. <input type="checkbox"/> Não <b>Passa para a questão 11</b>		
9. <input type="checkbox"/> Não sabe	<b>TRATGRAV</b>	<input type="checkbox"/>

47H. Tomou algum medicamento para induzir a ovulação?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

OVULA

48H. Fez inseminação artificial?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

INSEMINA

49H. Fez fertilização in vitro (bebê de proveta)?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

FERTILIZA

50H. Fez injeção de espermatozoides?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

INJECAO

#### **BLOCO I - CARACTERÍSTICAS DO PARTO E DO NASCIMENTO**

11. Qual foi o tipo de parto?

1.  Normal

2.  Cesárea **Passa para a questão 4I**

3.  Fórceps **Passa para a questão 4I**

4.  Vácuo extração **Passa para a questão 4I**

9.  Não sabe

TIPARTO

21. Se normal, a sra. fez episiotomia (corte, pique)?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

EPISIO

31. Quantas horas decorreram entre a internação e o parto normal? \_\_ \_

88.  Não se aplica

99.  Não sabe

HORASPARTO

41. Foi feita anestesia nas costas?

1.  Sim

2.  Não

9.  Não sabe

ANESTESIA

51. Caso cesárea, qual foi o motivo para fazer a cesárea?

1.  sofrimento fetal (batidas do coração do bebê diminuiu / ou o bebê fez cocô dentro da barriga da mãe)
2.  desproporção feto-pélvica (bacia pequena/bebê grande)
3.  distócia de apresentação (bebê sentado/ posição errada)
4.  hemorragia materna (teve sangramento)
5.  parada de progressão (parou trabalho de parto/ pararam as dores)
6.  eclâmpsia, pré-eclâmpsia (pressão alta)
7.  pós-maturidade (passou do tempo)
8.  morte fetal (o bebê morreu)
9.  diabetes materna (açúcar no sangue)
10.  cesáreas anteriores (já fez outra cesárea antes)
11.  laqueadura (para ligar trompas)
12.  mãe pediu (cesárea porque a mãe queria)
13.  médico quis (na hora o médico resolveu fazer cesárea)
14.  cesárea programada (médico marcou durante gravidez)
15.  cirurgias ginecológicas anteriores (Miomectomia, plástica perineal)
16.  outro \_\_\_\_\_

88.  não se aplica

99.  não sabe

MOTCESA

61. Quantas horas decorreram entre a internação e a cesárea?

--

88.  Não se aplica

99.  Não sabe

HORASCESA

71. A sra. ligou as trompas?

1.  Sim

2.  Não **Passe para a questão 91**

9.  Não sabe

LAQUEADURA

81. Qual o motivo pelo qual a sra. ligou as trompas?

1.  Já fez muita cesárea

2.  Por problemas de saúde. Qual? \_\_\_\_\_

3.  Questões financeiras

4.  Já tinha o número de filhos que desejava

5.  Outros \_\_\_\_\_

88.  Não se aplica

99.  Não sabe

MOTLAQ

99. O que a sra. sentiu que a fez vir para o hospital?

1.  Sangramento vaginal
2.  Perdeu líquido (água) vaginal
3.  Sentiu contração ou dor ou cólica/ barriga endurecida
4.  Febre/ infecção/infecção urinária
5.  O bebê parou de mexer/ diminuíram movimentos
6.  A vinda foi agendada para esta data
7.  O médico encaminhou
8.  Outro \_\_\_\_\_
99.  Não sabe

MOTHOSP

101. Caso tenha feito cesárea, quando a sra. internou já sabia que iria fazer cesárea?

1.  Sim
2.  Não
8.  Não se aplica
9.  Não sabe

SABIACESA

111. A sra. passou por outros serviços (maternidades) antes de vir para esse hospital?

1.  Sim
2.  Não **Passa para a questão 131**
9.  Não sabe

PEREGRINOU

121. Caso sim, por quantos serviços passou? \_

8.  Não se aplica
9.  Não sabe

QTPEREG

131. Quando a sra. foi hospitalizada estava sentindo as dores do trabalho de parto?

1.  Sim
2.  Não
9.  Não sabe

DORPARTO

141. O médico precisou romper a bolsa?

1.  Sim
2.  Não
9.  Não sabe

ROMPBOLSA

151. Foi preciso colocar soro ou outro remédio para começar o trabalho de parto ou para ajudar o bebê a nascer?

1.  Sim
2.  Não **Passa para a questão 181**
9.  Não sabe

SORONASC

161. Qual a medicação utilizada?

1.  Vaginal
2.  Soro (endovenosa)
8.  Não se aplica
9.  Não sabe

TIPOMED

171. Por que foi necessário ajudar o bebê a nascer?

1.  Passou do tempo de nascer
2.  Pressão alta
3.  Rompeu a bolsa
4.  Incompatibilidade sanguínea ( sangue não combina)
5.  O bebê estava morto
6.  O médico indicou
7.  O trabalho de parto parou
8.  Outra razão. Qual? \_\_\_\_\_

88.  Não se aplica

99.  Não sabe

MOTAJUDA

181. Quem atendeu ao parto?

1.  Médico
2.  Enfermeira
3.  Auxiliar de enfermagem
4.  Parteira leiga
5.  Outro
9.  Não sabe

QUEMPARTO

191. O parto foi realizado pelo mesmo médico que fez o pré-natal?

1.  Sim
2.  Não
9.  Não sabe

MEDICO

201. Qual a categoria de atendimento ao parto?

1.  SUS
2.  Plano de saúde/ seguro saúde
3.  Particular
9.  Não sabe

CATP



21. Na hora do nascimento, quem atendeu o RN na sala de parto?

1.  Médico obstetra
2.  Médico pediatra/neonatologista
3.  Anestesista
2.  Enfermeira
3.  Auxiliar de enfermagem
4.  Parteira leiga
5.  Outro \_\_\_\_\_
99.  Não sabe

ATENDRN

22. O pediatra falou com a sra na sala de parto antes ou depois que o bebê nasceu?

1.  Sim
2.  Não
9.  Não sabe

PEDIATRA

23. Número de filhos nascidos no parto: \_

9.  Não sabe

FETOS

#### **BLOCO J – EXPOSIÇÃO AO CITOMEGALOVÍRUS (CMV)**

1J. Na sua casa morou ou está morando alguma criança?

1.  Sim
2.  Não **Passa para a questão 8J**
9.  Não sabe

MORACRI

2J. Caso sim, ela tem até três anos de idade?

1.  Sim
2.  Não **Passa para a questão 8J**
8.  Não se aplica
9.  Não sabe

CRIMATE3

3J. Caso tenha até 3 anos de idade, ela morou com a sra. durante a gestação?

1.  Sim
2.  Não
8.  Não se aplica
9.  Não sabe

MOROUGEST

4J. Caso a criança não tenha morado com a sra. durante sua gestação, a criança morou com a sra. nos 12 meses anteriores?

1.  Sim
2.  Não
8.  Não se aplica
9.  Não sabe

MOROU12M

5J. Essa(s) criança(s) frequentam creches ou escolas?

1.  Sim

2.  Não **Passa para a questão 8J**

8.  Não se aplica

9.  Não sabe

CRECHE

6J. Quantos dias da semana a(s) criança(s) frequenta(m) a creche ou escola? \_

8.  Não se aplica

9.  Não sabe

DIASCRECHE

7J. Quantas horas por dia da semana a(s) criança(s) frequenta(m) a creche ou escola? \_\_

88.  Não se aplica

99.  Não sabe

HORACRECHE

8J. A sra. realizou algum cuidado direto com crianças de 3 anos ou menores (como trocar as fraldas, dar banhos, alimentar o bebê, etc.)? Incluindo crianças de sua família ou para amigos de seus filhos, ou no seu trabalho durante o ano que antecedeu ou atualmente na sua gravidez?

1.  Sim

2.  Não

8.  Não se aplica

9.  Não sabe

CUIDACRI

9J. Caso a sra. tenha realizado algum cuidado direto com crianças de 3 anos ou menores, qual a frequência de cuidados com essa criança?

1.  Menos de 12 vezes no ano

2.  Uma vez por mês

3.  Duas ou mais vezes no mês

4.  Uma vez por semana

5.  Mais de uma vez por semana

8.  Não se aplica

9.  Não sabe

QTCUIDA

### **BLOCO K – DADOS DO PRONTUÁRIO**

1K. Há registro de administração de ocitocina durante o trabalho de parto?

1.  Sim

2.  Não

OCITOCINA

2K. Horário de início do medicamento (indução): \_\_: \_\_

8888.  Não se aplica 9999.  Não informado

HORAIMED

3K. Horário do término do medicamento (indução): \_\_: \_\_

8888.  Não se aplica 9999.  Não informado

HORATMED

4K. Caso o parto tenha sido cesárea, anotar a indicação da cesárea do prontuário \_\_\_\_\_

88.  Não se aplica 99.  Não informado

INDICACESA

5K. Registro da idade gestacional avaliada pela Ultrassonografia (anotar o primeiro ultrassom) \_\_ semanas

99.  Não informado

IDGESTUS

6K. Data da Ultrassonografia \_\_/\_\_/\_\_\_\_

99999999.  Não informado

DATAUS

## ANEXO D – Questionário do RN



## QUESTIONÁRIO DO NASCIMENTO - RN

**BLOCO A – DADOS DE IDENTIFICAÇÃO**

1A. Número de identificação: -----

- 1ª casela: 1 Ribeirão Preto  
2 São Luís

- 2ª casela: 1 Coorte iniciada no Pré-natal  
2 Coorte iniciada no nascimento

- 3ª casela: 1 Questionário do pré-natal  
2 Questionário do nascimento  
3 Questionário do 1º ano

- 4ª casela: 0 Questionário do nascimento-mãe  
1 Questionário do nascimento RN 1  
2 Questionário do nascimento RN 2  
3 Questionário do nascimento RN 3  
4 Questionário do nascimento RN 4  
5 Questionário do nascimento RN 5  
8 saliva  
9 cordão

5ª à 8ª. casela: número sequencial para cada cidade

2A. Cidade:

1.  Ribeirão Preto  
2.  São Luís

3A. Data da Entrevista (DD/MM/AAAA):

--/--/----

4A. Entrevistador (a):

5A. Data do nascimento do RN (DD/MM/AAAA):

--/--/----

6A. Caso tenha nascido mais de um filho, indique qual a sequência de nascimento desse RN?

1.  1º  
2.  2º  
3.  3º  
4.  4º  
5.  5º  
8.  Não se aplica  
9.  Não sabe

7A. Qual o sexo do recém-nascido?

1.  Masculino  
2.  Feminino  
9.  Não sabe

8A. Como nasceu o filho da sra.?

1.  Vivo  
2.  Morto  
9.  Não sabe

NUMERO

CIDADE

DATAENT

DATANASC

ORDEMNASC

SEXO

NASC

9A. Qual foi a apresentação do bebê na hora do parto?

- 1.  Cefálica
- 2.  Pélvica
- 3.  Transversa
- 9.  Não sabe

APRESENT

**Caso o RN tenha nascido morto, passe para o Bloco B**

10A. Qual nome a sra. pretender dar ao seu bebê?

\_\_\_\_\_

NOME

11A. O RN apresentou algum problema de saúde?  
(perguntar para a mãe e anotar do prontuário)

- 1.  Sim
- 2.  Não **Passe para a questão 17A**
- 8.  Não se aplica
- 9.  Não sabe

PROBLEMA

Caso tenha apresentado problema, nós diga qual foi.

12A. Problema

- 88.  Não se aplica
- 99.  Não sabe

PROB1

13A. Problema

- 88.  Não se aplica
- 99.  Não sabe

PROB2

14A. Problema

- 88.  Não se aplica
- 99.  Não sabe

PROB3

15A. Problema

- 88.  Não se aplica
- 99.  Não sabe

PROB4

16A. Problema

- 88.  Não se aplica
- 99.  Não sabe

PROB5

17A. Após o nascimento, quando saiu da sala de parto, o RN foi para:

1.  Berço ao lado da mãe
2.  Berçário
3.  Cama da mãe
4.  UTI Neonatal
5.  Outro \_\_\_\_\_
8.  Não se aplica
9.  Não sabe

LOCALRN

### BLOCO B – DADOS DO PRONTUÁRIO

1B. Horário de nascimento do RN: \_\_\_\_:\_\_\_\_

HORANASC

2B. Apgar 1º minuto: \_\_

99.  Não avaliado

APGAR1

3B. Apgar 5º minuto: \_\_

99.  Não avaliado

APGAR5

4B. Em caso de nascido morto:

1.  Antes do parto
2.  No momento do parto
8.  Não se aplica
9.  Não sabe

TIPONASC

5B. O RN apresentou algum defeito congênito (malformação)?

1.  Sim
2.  Não Passe para a questão 7B
9.  Não sabe

DEFCONG

6B. Qual foi o tipo de defeito apresentado pelo RN?

1.  Gastrosquise
2.  Onfalocele
3.  Defeitos do tubo neural - meningomielocoele
4.  Defeitos do tubo neural - encefalocele
5.  Trato genito-urinário - unilateral
5.  Trato genito-urinário - bilateral
6.  Cardíaco. Qual? \_\_\_\_\_
7.  Outro \_\_\_\_\_
88.  Não se aplica
99.  Não sabe

TIPODEFCONG

**ANTROPOMETRIA DO RN**

7B. Peso do RN: \_ . \_ . \_ . \_ g

9999.  Não avaliado

PESONASC

8B. Comprimento: \_ . \_ . \_ cm

999.  Não avaliado

COMPNASC

9B. Perímetro cefálico: \_ . \_ . \_ cm

999.  Não avaliado

PCNASC

10B. Peso da placenta: \_ . \_ . \_ . \_ kg

9999.  Não avaliado

PLACENTA

**MATERIAL BIOLÓGICO DO RN**

11B. Foi feita coleta de tecido do cordão umbilical?

1.  Sim2.  Não

CORDAO

12B. Foi feita coleta de saliva?

1.  Sim2.  Não

SALIVA